

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 09:54:32 ; Search time 89.59 Seconds
(without alignments)
332.928 Million cell updates/sec

Title: US-09-668-482-2
Perfect score: 2563
Sequence: 1 MGLYTLMTVFLCTIVLPVLL.....GPTIYPVDNLPTRKTSYVRN 492

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 segs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601: *
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT: *
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT: *
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT: *
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT: *
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT: *
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT: *
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT: *
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT: *
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT: *
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT: *
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT: *
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT: *
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT: *
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT: *
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT: *
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT: *
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT: *
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT: *
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT: *
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT: *
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT: *
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

8					
Result No.	Score	Query Match	Length DB	ID	Description
1	2563	100.0	492	19 AAW37733	Cytochrome zP450RA
2	2563	100.0	492	19 AAW44159	zebrafish retinoid
3	1738	67.8	497	19 AAW37734	Human cytochrome P
4	1738	67.8	497	19 AAW44160	Human retinoid met
5	1734	67.7	497	19 AAW37735	Cytochrome P450RAI
6	1734	67.7	497	19 AAW44161	Mouse retinoid met
7	752.5	29.4	216	21 AAB12489	Human PSEC64 prote
8	552.5	21.6	472	18 AAW27153	Arabidopsis thalia
9	552.5	21.6	472	21 AAG44571	Arabidopsis thalia
10	552.5	21.6	472	21 AAG45022	Arabidopsis thalia
11	552.5	21.6	491	21 AAG45021	Arabidopsis thalia

12	552.5	21.6	492	21 AAG44570	Arabidopsis thalia
13	536	20.9	444	21 AAG44572	Arabidopsis thalia
14	536	20.9	444	21 AAG45023	Arabidopsis thalia
15	503	19.6	513	21 AAB07921	A cytochrome P450
16	481.5	18.8	461	21 AAG46490	Arabidopsis thalia
17	481.5	18.8	462	21 AAG46489	Arabidopsis thalia
18	481.5	18.8	465	21 AAG46491	Arabidopsis thalia
19	475.5	18.6	461	21 AAG11836	Arabidopsis thalia
20	475.5	18.6	462	21 AAG11835	Arabidopsis thalia
21	475.5	18.6	465	21 AAG11834	Arabidopsis thalia
22	470	18.3	468	21 AAG20783	Arabidopsis thalia
23	425.5	16.6	430	21 AAG20784	Arabidopsis thalia
24	397	15.5	163	21 AAB40557	Human ORFX ORF321
25	394.5	15.4	489	21 AAG30048	Arabidopsis thalia
26	394	15.4	481	21 AAG30049	Arabidopsis thalia
27	366.5	14.3	433	21 AAG30050	Arabidopsis thalia
28	366	14.3	475	21 AAB07676	Amino acid sequenc
29	343	13.4	511	20 AAY23341	A P450-2 protein i
30	343	13.4	511	21 AAB19694	Sweetgum conferyl
31	342	13.3	471	21 AAG33013	Arabidopsis thalia
32	342	13.3	479	21 AAG33012	Arabidopsis thalia
33	339.5	13.2	490	17 AAR89862	Cytochrome P450 2C
34	339.5	13.2	490	17 AAR89865	Cytochrome P450 2C
35	339.5	13.2	490	19 AAW64070	Human cytochrome P
36	339.5	13.2	490	19 AAW64073	Human cytochrome P
37	339.5	13.2	490	20 AAY04127	Mammalian cytochro
38	339	13.2	469	21 AAG23014	Arabidopsis thalia
39	333	13.0	388	21 AAG20785	Arabidopsis thalia
40	332	13.0	490	16 AAR72361	Human cytochrome P
41	332	13.0	490	17 AAR93168	Human cytochrome P
42	332	13.0	490	17 AAR81465	Human derived cyto
43	332	13.0	513	22 AAB31008	Amino acid sequenc
44	332	13.0	513	22 AAB48181	B. napus F5H polyp
45	330.5	12.9	520	18 AAW26640	Arabidopsis thalia

ALIGNMENTS

RESULT	1				
AAW37733					
ID	AAW37733	standard;	Protein;	492	AA.
XX					
AC	AAW37733;				
XX					
DT	07-JUL-1998	(first entry)			
XX					
DE	Cytochrome zP450RAI	protein.			
XX					
KW	Retinoid regulated gene;	cytochrome P450 gene;	enzyme;		
KW	oxidative metabolism;	P450RAI; retinoic acid;	RA; promoter.		
XX					
OS	Danio rerio.				
XX					
PN	WO9749832-A2.				
XX					
PD	31-DEC-1997.				
XX					
PF	23-JUN-1997;	97WO-CA00488.			
XX					
PR	01-OCT-1996;	96US-0724466.			
XX	21-JUN-1996;	96US-0667546.			
PA	(TOOH) UNIV QUEENS KINGSTON.				
XX					
PI	Petkovich PM;				
XX					
DR	WPI; 1998-077193/07.				
XX	N-PSDB; AAV09251.				
PT					
PT	Identifying DNA encoding inducible or suppressible cytochrome P450 -				
PT	by screening for drugs which reduce the catabolism of retinoic acid,				
PT	useful in cancer chemotherapy and the treatment of acne and				

PT psoriasis
XX
PS Example 1; Pages 53-55; 113pp; English.
XX
CC This is the amino acid for cytochrome zp450RAI of the zebra fish.
CC Its expression is dependent on the presence of retinoic acid (RA).
CC The retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RAI is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen using
CC promoters and nucleotide sequences to identify drugs which are useful
CC for reducing the catabolism of RA.
XX
SQ Sequence 492 AA;

Query Match 100.0%; Score 2563; DB 19; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.3e-227;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLYTLMTFLCTIVLPVLLFLAAVKLWEMLMIRVDPNCRSPPLPGTMGLPFIGETLQL 60
DB 1 mgllytlmvtflctivlpvllflaavklwemlmirrvdpncrspplpgtmglpfigetlql 60
QY 61 ILQRRKFLRMKROKYGCIYKTHLFGNPTVRVMGADNVROILLGEHKLVSQWPAASVRTIL 120
DB 61 ilqrkflrmkrqkygciythlfgnptvrvmgadnvrqillgehklvsvqwpasvrtil 120
QY 121 GSDTLSNHGVQHKNNKKKAIMRAFSRDALAHYIPVIOQEVKSAIQEWLQKDSCLVYPPEM 180
DB 121 gsdtlsnhgvqhknkkkaimrafsrdalehyipviqevksaigewlqkdsclvypem 180
QY 121 gsdtlsnhgvqhknkkkaimrafsrdalehyipviqevksaigewlqkdsclvypem 180
DB 121 gsdtlsnhgvqhknkkkaimrafsrdalehyipviqevksaigewlqkdsclvypem 180
QY 181 KKLMEIRIAMRILLGFEPEQIKTDEQELVEAFEEEMIKNLFSLPIDVPFSGLYRGLRARNF 240
DB 181 kklmfriamrillgfepeqiktdeqelveafeemiknlfslpidvpfsglyrqlrar nfi 240
QY 241 HSKIEENIRKKIQDDNENEQKYKDALQLLIENSRRSDEPFSLOAMKEATELLFGGHET 300
DB 241 hskieenirkkIQDDNENEQKYKDALQLLIENSRRSDEPFSLOAMKEATEllfgghet 300
QY 301 TASTATSLVMEFLGNTDEVVQKRVREVEQKVMGYTPGKGLSMELLDQLKTYTCVIKETL 360
DB 301 tastatslvmeflgntdevvqkvrveeqkvmgytpgkglsmelldqlktytcviketl 360
QY 361 RINPVPVGGFRVALKTFELNGYQIPKGNVYISICDTHDVADVFPNKEEFQPERFMSKGL 420
DB 361 rinpvpvggfrvalktfelngyqipkgnvvisicdthdvadvfpnkeefqperfmskgl 420
QY 421 EDGSRFNYIPFGGSRMCVGKEFAKVLKIFLVELTQHCHNWILSNGPPTMKGTPTIYPVD 480
DB 421 edgsrfnyipfggssrmcvgkefakvlkiflveltqhc nwlsgpptomktgtptiypvd 480
QY 481 NLPTKFTSYVRN 492
DB 481 nlptkftsyrn 492

RESULT 2
AAW44159
ID AAW44159 standard; Protein; 492 AA.
XX
AC AAW44159;
XX
DT 22-JUN-1998 (first entry)
XX
DE zebrafish retinoid metabolising protein zp450RAI.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; zebrafish; inhibitor; antibody; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;

KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening.
XX
OS Danio rerio.
XX
PN WO9749815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI; 1998-077178/07.
DR N-PSDB; AAV12203.
XX
PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 1; Page 50-51; 110pp; English.
XX
CC This protein comprises a novel zebrafish retinoid metabolising
CC protein, designated zp450RAI. Its amino acid sequence was deduced
CC from a cDNA clone (see AAV12203) isolated from a 6-18 hr embryo
CC library. It includes a haem-binding motif characteristic of
CC cytochrome P450 proteins. zp450RAI is a retinoid oxidase that has
CC the ability to hydroxylate retinoic acid at the 4 position of the
CC beta-ionone ring, and is inducible in epithelial cells exposed to
CC retinoic acid. Zebrafish, human and mouse P450RAIs (see AAW44159-61)
CC are claimed. They can be expressed in host cells and used to
CC metabolize retinoic acid in an organism or cell, in drug screening,
CC and to raise antibodies useful for inhibiting retinoic acid
CC hydroxylation for the treatment of cancer, actinic keratosis, oral
CC leukoplakia, secondary tumours of the head and/or neck, non-small
CC cell lung carcinomas, basal cell carcinomas, acute promyelocytic
CC leukaemia, skin cancer, and premalignancy associated with actinic
CC keratosis, acne, psoriasis and/or ichthyosis.
XX
SQ Sequence 492 AA;

Query Match 100.0%; Score 2563; DB 19; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.3e-227;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLYTLMTFLCTIVLPVLLFLAAVKLWEMLMIRVDPNCRSPPLPGTMGLPFIGETLQL 60
DB 1 mgllytlmvtflctivlpvllflaavklwemlmirrvdpncrspplpgtmglpfigetlql 60
QY 61 ILQRRKFLRMKROKYGCIYKTHLFGNPTVRVMGADNVROILLGEHKLVSQWPAASVRTIL 120
DB 61 ilqrkflrmkrqkygciythlfgnptvrvmgadnvrqillgehklvsvqwpasvrtil 120
QY 121 GSDTLSNHGVQHKNNKKKAIMRAFSRDALAHYIPVIOQEVKSAIQEWLQKDSCLVYPPEM 180
DB 121 gsdtlsnhgvqhknkkkaimrafsrdalehyipviqevksaigewlqkdsclvypem 180
QY 181 KKLMEIRIAMRILLGFEPEQIKTDEQELVEAFEEEMIKNLFSLPIDVPFSGLYRGLRARNF 240
DB 181 kklmfriamrillgfepeqiktdeqelveafeemiknlfslpidvpfsglyrqlrar nfi 240
QY 241 HSKIEENIRKKIQDDNENEQKYKDALQLLIENSRRSDEPFSLOAMKEATELLFGGHET 300
DB 241 hskieenirkkIQDDNENEQKYKDALQLLIENSRRSDEPFSLOAMKEATEllfgghet 300
QY 301 TASTATSLVMEFLGNTDEVVQKRVREVEQKVMGYTPGKGLSMELLDQLKTYTCVIKETL 360

Db 301 laatastslvmflglntevvqkvreevqekvmgmtytpgkylsmelldqlytgcvtketl 360
QY 361 RINPPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVADEVFNKKEEFQPERFMSKGL 420
Db 361 rinppvpggfrvalktfelngyqipkgwnviysicdthdvadvfnakeefqperfmskgl 420
QY 421 EDGSRPNYIPFGGSRMCGKEFAKVLKIFLVELTQHCHNWLNSNGPPTMKTGPTIYPVD 480
Db 421 edgsrpnypifggsgsmcvgkefakvllkiflveltqhcnwlinsngpptmktgptiypvd 480
QY 481 NLPFKFTSYVRN 492
Db 481 nlpcfkftsyrn 492

RESULT 3
AAW37734
ID AAW37734 standard; Protein: 497 AA.

AC AAW37734;
XX 07-JUL-1998 (first entry)
DE Human cytochrome P450RAI protein.
XX Retinoid regulated gene; cytochrome P450 gene; enzyme;
KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter.
XX Homo sapiens.
XX WO9749832-A2.
XX 31-DEC-1997.
XX 23-JUN-1997; 97WO-CA00488.
XX 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX (TOOH) UNIV QUEENS KINGSTON.
XX Petkovich PM;
PI WPI; 1998-077193/07.
DR N-PSDB; AAV09247.
XX

PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis

PS Example 4; Pages 58-59; 113pp; English.

XX This is the amino acid sequence of the human cytochrome P450RAI. Its
CC expression is dependent on the presence of retinoic acid (RA). The
CC retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RAI is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen
CC using promoters and nucleotide sequences to identify drugs which are
CC useful for reducing the catabolism of RA.

XX Sequence 497 AA;

Query Match 67.8%; Score 1738; DB 19; Length 497;
Best Local Similarity 68.0%; Pred. No. 3.6e-151;
Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;

QY 1 MGLYTLMTFLCTIVLPVLLFLAAVKLMEMLMIRVDPNCRSPDPRTMGLPFTEGLQL 60
Db 1 mglpallasaalcftvlpillflaalklwdlycvsgdrscslpippgtmgfifgetlqm 60
QY 61 ILQRRKFLRMKROKYGCIYKTHLFGNPTVRVMGADNVKQILGHEKLVSVQVWPASVYRIL 120
Db 61 vlgrrkflgmkrkygfikthlfgprlvrmgadnvrillgddrlsvshvpasvrtll 120
QY 121 GSDTLSNVHGVQHNKKKAIMRAFSRPALEHYIPVIOQEVKSAIQEWLQKDSK---VLV 176
Db 121 gsgclsnldshkqrrkvimrafsralecypviteevgssleqwl---scgergliv 177
QY 177 YPEMKKLMERIAMRILGFEPQIKTD---EQELVEAFEEEMIKNLFSLPIDVPSGLYRG 233
Db 178 ypevkrilmfiamrillgcep-qlagdgsseqqlveafeemtrnlfsldvpsgllyrg 236
QY 234 LRARNFIHSKIEENIRKKIOD-DDNENQKYKDALQLIENSRSDPEPSLOAKREATE 292
Db 237 mkarnlharlegulrakicglraseagqgckdalqllhswergerldmqalkqsste 296
QY 293 LLEGGHETTASTATSLVFLGLNTEVVQKVRREVQEKVEMGMYTPGKGLSMELDLQKYT 352
Db 297 llfghettasatslitylglphvhlqkvrleelkskgllicksnqdnklldmellqlyi 356
QY 353 GCVIKETLRINPPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVADEVFNKKEFQ 412
Db 357 gcvikektrlnppvpggfrvalktfelngyqipkgwnviysicdthdvaefnkeefnp 416
QY 413 ERFMSKLEDDGSRPNYIPFGGSRMCGKEFAKVLKIFLVELTQHCHNWLNSNGPPTMKT 472
Db 417 drfsaphedastisifipfggslrscvgykefakillkiflvelarhcdwqlngpptmkt 476
QY 473 GPTIYPVDNLPTKFTSY 489
Db 477 sptlvypvdnlparthf 493

RESULT 4
AAW44160
ID AAW44160 standard; Protein: 497 AA.

XX AAW44160;
AC 22-JUN-1998 (first entry)
DE Human retinoid metabolising protein hp450RAI.

XX Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; human; inhibitor; antibody; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening.

XX Homo sapiens.

XX WO9749815-A1.

XX 31-DEC-1997.

XX 23-JUN-1997; 97WO-CA00440.

XX 01-OCT-1996; 96US-0724466.

PR 21-JUN-1996; 96US-0667546.

XX (TOOH) UNIV QUEENS KINGSTON.

XX Beckett BR, Jones G, Petkovich PM, White JA;

DR WPI; 1998-077178/07.

DR N-PSDB; AAV12204.

XX Retinoid metabolising protein - useful to develop products to treat,

PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis

PS Claim 1, Page 54-55; 110pp; English.

CC This protein comprises a novel human retinoid metabolising protein,
CC designated hp450RAI. Its amino acid sequence was deduced from a
CC cDNA clone (see AAV12204) isolated from a retinoic acid-treated NT2
CC cell library. It includes a haem-binding motif characteristic of
CC cytochrome P450 proteins. hp450RAI is a retinoid oxidase that has
CC the ability to hydroxylate retinoic acid at the 4 position of the
CC beta-ionone ring, and is inducible in epithelial cells exposed to
CC retinoic acid. Zebrafish, human and mouse P450RAIs (see AAM44159-61)
CC are claimed. They can be expressed in host cells and used to
CC metabolize retinoic acid in an organism or cell, in drug screening,
CC and to raise antibodies useful for inhibiting retinoic acid
CC hydroxylation for the treatment of cancer, actinic keratosis, oral
CC leukoplakia, secondary tumours of the head and/or neck, non-small
CC cell lung carcinomas, basal cell carcinomas, acute promyelocytic
CC leukaemia, skin cancer, and premalignancy associated with actinic
CC keratosis, acne, psoriasis and/or ichthyosis.

XX Sequence 497 AA;

Query Match

Best Local Similarity 67.8%; Score 1738; DB 19; Length 497;
Matches 338; Conservativity 63; Mismatches 84; Indels 12; Gaps 5;

QY 1 MGLYTLMTWTECTIVLPVLLFLAAYKIMEMLMIRVDNCRSPPLPGTMGLPFTGTLQL 60
DB 1 mglpalalasalctfvlplllflaalklwdlycvsgdrscalplppgtmgfifggetlqm 60
QY 61 ILQRRKFLRMKROKYGCITYKTHLFGNPTVRVMGADNVROILLGEHKLVSQWPAVYRTIL 120
DB 61 vlqrrkflgmkrkygfyikthlfgprtvrmgadnvrtilldgdrllsvshwpaavrtll 120
QY 121 GSDTLNVHGVQHNKKKAIMRAFSRDALHYIPVIOEYKSAIQEWLQKDSCLV 176
DB 121 gsgclsnldshskgrkkmratfsrealcecyvpteevsgslqwl--scgergliv 177
QY 177 YPEMKKLMFRIAMRILLGFEPEQIKTD--EQELVEAFEFEMIKNLFSLPIDVPFSGLYRG 233
DB 178 ypevkrimfriamrillgcep-qiaqdgdsqqlveafeemtrnlfsldvpfsglyrg 236
QY 234 LRARNFHSKIEENIRKKIQD-DNENEQKYKDALLIENSRSDPEFSLQAMKEATE 292
DB 237 mkarnliharieenirakicgliraseagqgckdalqllshswgerldmgalkgsste 296
QY 293 LIFGHEHTASTATSLVMFLGLNTEVQKRVREVOEKVEMGMVTPGKGLSMELLDQLKYT 352
DB 297 lfghettasatstltyglpvhlyqkvreelskgljcksnqndklmelleqlky 356
QY 353 GCVTKETLRINPPVPGGFRAVALKTEELNGYQIPKGMNVIYSICDTHDVADVPNKEEQP 412
DB 357 gcvtketlrinppvpggfrralktfelngyqipkgwnvisicdthdvaefnkeefn 416
QY 413 ERFMSKGLDEGSRFNYPFGGSSRMCVGKEFAKVLKTFVVELTQHCHNWLNSGPPMTKT 472
DB 417 drfsaphgedasrfsfipfggslrscvqgkefakllkiftvelarhcdwqlngppmtkt 476
QY 473 GPTIYVVDNLPTKFTSY 489
DB 477 sptcyvvdnlparfthf 493

RESULT 5

AAW37735 standard; Protein; 497 AA.

AC AAW37735;
XX 07-JUL-1998 (first entry)
DT

XX Cytochrome P450RAI isoform.

KW Retinoid regulated gene; cytochrome P450 gene; enzyme;
KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter; isoform.

OS Synthetic.

PN WO9749832-A2.

PD 31-DEC-1997.

PF 23-JUN-1997; 97WO-CA00488.

PR 01-OCT-1996; 96US-0724466.

PR 21-JUN-1996; 96US-0667546.

PA (T00H) UNIV QUEEN'S KINGSTON.

PI Petkovich PM;

DR WPI; 1998-077193/07.

DR N-PSDB; AAV09252.

PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis

XX Disclosure; Pages 596-59H; 113pp; English.

CC This amino acid sequence is of an isoform of cytochrome zp450RAI.
CC Its expression is dependent on the presence of retinoic acid (RA).
CC The retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome P450
CC nucleotide sequence can be used to induce or suppress the expression
CC of its protein. P450RAI is highly induced by RA in cell lines and
CC tissues. This allows for development of a drug screen using promoters
CC and nucleotide sequences to identify drugs which are useful for
CC reducing the catabolism of RA.

XX Sequence 497 AA;

Query Match

Best Local Similarity 67.7%; Score 1734; DB 19; Length 497;
Matches 336; Conservativity 61; Mismatches 89; Indels 10; Gaps 4;

QY 1 MGLYTLMTWTECTIVLPVLLFLAAYKIMEMLMIRVDNCRSPPLPGTMGLPFTGTLQL 60
DB 1 mglpalalasalctfvlplllflaalklwdlycvsgdrscalplppgtmgfifggetlqm 60
QY 61 ILQRRKFLRMKROKYGCITYKTHLFGNPTVRVMGADNVROILLGEHKLVSQWPAVYRTIL 120
DB 61 vlqrrkflgmkrkygfyikthlfgprtvrmgadnvrtilldgdrllsvshwpaavrtll 120
QY 121 GSDTLNVHGVQHNKKKAIMRAFSRDALHYIPVIOEYKSAIQEWLQKDSCLV 176
DB 121 gsgclsnldshskgrkkmratfsrealcecyvpteevsgslqwl--scgergliv 177
QY 177 YPEMKKLMFRIAMRILLGFEPEQIKTD--EQELVEAFEFEMIKNLFSLPIDVPFSGLYRG 234
DB 178 ypevkrimfriamrillgceppaggsedeqqlveafeemtrnlfsldvpfsglyrgv 237
QY 235 RARNFHSKIEENIRKKIQD-DNENEQKYKDALLIENSRSDPEFSLQAMKEATEL 293
DB 238 karnliharieenirakirrlqatepdgckdalqllshswgerldmgalkgsste 297
QY 294 LFGHEHTASTATSLVMFLGLNTEVQKRVREVOEKVEMGMVTPGKGLSMELLDQLKYT 353
DB 298 lfghettasatstltyglpvhlyqkvreelskgljcksnqndklmelleqlkyig 357

OY	354	CVIKETLRINPVPVGGFRVALKTFELNGYOIPKGWNVIYSICDTHDVADVFPNKEEFOPE	413
		: : : :	
Dd	358	cviKetlrInppvpgfrvalktfelngyqipkgwnvisctdhvadiftnkeefnpd	417
OY	414	FEMSKGLEDSRENYIPEGGSRMCVCKEFVKLLKIPLVELTQHCHNWILSNGPPTMKTC	473
		: : :	
Dd	418	rfiVhpDeasfStfiPgglrscvgkEfakIlkliftVelarHcdwqlIngptmKts	477
OY	474	PtiYPVDNLPIKFtSY	489
		: : :	
Dd	478	pTiypvdnlparitfyf	493

RESULT	6
AAW44161	
ID	AAW44161 standard; Protein; 497 AA.
XX	
AC	AAW44161;
XX	
DT	22-JUN-1998 (first entry)
XX	
DE	Mouse retinoid metabolising protein mp450RAI.
XX	
KW	Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW	retinoic acid; mouse; inhibitor; antibody; cancer;
KW	actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW	non-small cell lung carcinoma; basal cell carcinoma;
KW	acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW	ichthyosis; therapy; diagnosis; screening.
XX	
OS	Mus musculus.
XX	
PN	WO9749815-A1.
XX	
PD	31-DEC-1997.
XX	
PF	23-JUN-1997; 97WO-CA00440.
XX	
PR	01-OCT-1996; 96US-0724466.
PR	21-JUN-1996; 96US-0667546.
XX	
PA	(TOOH) UNIV QUEENS KINGSTON.
XX	
PI	Beckett BR, Jones G, Petkovich PM, White JA;
XX	
DR	WPI; 1998-071718/07.
DR	N-PSSB; AAV12205.
XX	
PT	Retinoid metabolising protein - useful to develop products to treat,
PT	e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT	ichthyosis
XX	
PS	Claim 1; Page 65-66; 110pp; English.
XX	
CC	This protein comprises a novel mouse retinoid metabolising protein,
CC	designated mp450RAI. Its amino acid sequence was deduced from a
CC	cDNA clone (see AAV12205) isolated from a retinoic acid-treated P19
CC	teratocarcinoma library. It includes a haem-binding motif
CC	characteristic of cytochrome P450 proteins. mp450RAI is a retinoid
CC	oxidase that has the ability to hydroxylate retinoic acid at the 4
CC	position of the beta-lonone ring, and is inducible in epithelial
CC	cells exposed to retinoic acid. Zebrafish, human and mouse P450RAIs
CC	(see AAW44159-61) are claimed. They can be expressed in host cells
CC	and used to metabolize retinoic acid in an organism or cell, in drug
CC	screening, and to raise antibodies useful for inhibiting retinoic
CC	acid hydroxylation for the treatment of cancer, actinic keratosis,
CC	oral leukoplakia, secondary tumours of the head and/or neck,
CC	non-small cell lung carcinomas, basal cell carcinomas, acute
CC	promyelocytic leukaemia, skin cancer, and premalignancy associated
CC	with actinic keratosis, acne, psoriasis and/or ichthyosis.
XX	
SQ	Sequence 497 AA;

Query Match	67.7%;	Score 1734;	DB 19;	Length 497;
Best Local Similarity	67.7%;	Pred. No. 8.4e-151;		
Matches 336;	Conservative 61;	Mismatches 89;	Indels 10;	Gaps 4;

QY		1	MGLYTLMWTFCLCTIVLPVLLFLAAVKIMEMLMIRVDPNCRSDPLPRTMGLPFIGETLQL	60
Dd		1	mglpalasalcufvlpillflaaiklwdlycvssrdsrscalpibp9tmgfiffigetlqm	60
QY		61	ILQRKFLRMKRKYGCITYKTHLFGNPTVRVMGADNRQILLGEHKLVSVOWPASVRTIL	120
Dd		61	vlgrrkfllgmkrxkygfikycthlfgrptvrvmgadnvrtiligehrlvsvhwpasvrtil	120
QY		121	GSDTLSNHVGVQHKKKKKAIMRAFSPDLEHYIPVIOQEVSALIOEWLOKDSC---	VLV 176
Dd		121	gagclsnlhdsnhqrkfkvimqafsréalqcyvlviaeavssclegwl---scgeygliv	177
QY		177	YPEMKKLMFRIAMRILLGFEPED--KTDEQELVEAFEMEKNLFSLPIDVPFGSHYRG	L 234
Dd		178	ypevkrlmfriamrillgccepgpaggsedeqlveafeemtrnlfsldvpfgshyrgv	237
QY		235	RARNFIHSKIEENTRRKKIOD-DDNENEQKYKDALLQILIENSRRSDEPFSLOAMKEATEL	293
Dd		238	karnliharieenlrakirrlgatepdgckdalqllihsawgerlamgakgstel	297
QY		294	LEGGHETTASTATSLMFELGLNTEVVQKVREEVQEKVENMGMTTPCKGLSMELLDOJLYTG	353
Dd		298	lfghettasaatslityglphvhlqvreeiksksyllcksngdnklametleqklyig	357
QY		354	CVIKETLRINPPVPGGFRVALKTFELNGYQIPKGWNVIYSICDTHDVADVFPNKERPOPE	413
Dd		358	cviketlrinppvggfrvalktfelngyqipkgwnviysicdhvadiftнкеerfpd	417
QY		414	RFMSKGLEDSRRNYIPFGGSRMCVCFEFAVLKIFLVELTQHCHNWILSNGRPMTKG	473
Dd		418	rfiwphedastrfipfgglsrscvgkefakilkiiftvelarhcdwqlnngprmtks	477
QY		474	PTIYPVDNLPTKFTSY 489	
Dd		478	ptvyvdnlparfyf 493	

RESULT	7
AAB12489	
ID	AAB12489 standard; Protein; 216 AA.
XX	
AC	AAB12489;
XX	
DT	31-OCT-2000 (first entry)
XX	
DE	Human PSEC64 protein sequence SEQ ID NO:2.
XX	
KW	Human; PSEC64; neuron growth; nerve disease.
XX	
OS	Homo sapiens.
XX	
PN	JP2000152790-A.
XX	
PD	06-JUN-2000.
XX	
PF	19-NOV-1998; 98JP-0329989.
XX	
PR	19-NOV-1998; 98JP-0329989.
XX	
PA	(HERT-) HIRIKKUSU KENKYUSHO KK.
XX	
DR	WPI; 2000-468126/41.
DR	N-PSDB; AAA60752.
XX	
PT	A protein related to the growth of neuron and a gene encoding said
PT	protein -
XX	

PS Claim 1; Page 10; 13pp; Japanese.

CC The present sequence represents a human protein, designated PSEC64, which
CC is related to neuron growth. The PSEC64 protein and its gene can be used
CC for the development of a preventive agent for use in the treatment of
CC diseases in which nerves are involved.

SQ Sequence 216 AA;

Query Match	29.48;	Score 752.5;	DB 21;	Length 216;
Best Local Similarity	68.28;	Pred. No. 3.9e-61;		
Matches 150; Conservative	28;	Mismatches 31;	Indels 11;	Gaps 4;

```

OY 1 MGBYTLMTWFLCTIVLPVLLFLAANKLWMLMIRVDPNCRBPLPGTMGLPIGEITL 60
    ||| | : ||| ||| ||| ||| ||| : | : ||| ||| ||| ||| ||| :
Db 1 mglpallasalctfviplllflaalkwldlycvsgrdrscslpibpqtmgfpffgetlqm 60

OY 61 ILORRRKFLRMKROKYYGCIYKTHLFNPTRVVMGADNVRQILGHEKLVSVQWPPASVRTL 1200
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 vlgrtkfllgmkrkygflykthlfgrptlvrmgadnvrlllgehrlsvhmpasvrtl 1200

OY 121 GSDTLNVNHGVQHKNNKKKAIMRAFSDALEHYIPVIOQEVKSAIOEWLQKDSC---VLV 1766
    ||| ||| : || : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 gsgclsnldshshkqrrkvlmrafsrealecyvpvltteavsgslqwl---scgergliv 1777

OY 177 YPEMKKLMFRIAMRILGFEPEQIKTD---EQELVEAFEE 213
    ||| : ||| ||| ||| ||| ||| ||| : | : ||| ||| ||| ||| ||| |||
Db 178 ypevkrllmfriamrilllgcep-qlagdgdsesqglveafee 216

```

RESULT	8
AAW27153	
ID	AAW27153 standard; protein; 472 AA

DT	14-APR-1998	(first entry)
XX		
AC	AAW27153;	
...		

DE *Arabidopsis thaliana* cytochrome P450-type hydroxylase.

KW Cytochrome P450-type hydroxylase; identification; brassinosteroid;
KW brassinosteroid inhibitor; modified plant; recombinant production;
KW teasterone.

OS *Arabidopsis thaliana*.

PN W09735986-A1

PD 02-OCT-1997

PF 27-MAR-1997; 97WO-EP01586.

PR 27-MAR-1996; 96US-0622166.

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN

Altman T, Koncz C, Mathur J, Szekeres MA: PI

DR WPI; 1997-489649/45.

DN FDD, 44100000, 44100000/

PT New isolated plant cytochrome P450-type hydroxylase gene - used to
PT identify substances acting as brassino-steroid(s) or brassinosteroid
PT inhibitors for the production of modified plants

PS Claim 1; Pages 44-46; 77pp; English

The present sequence is *Arabidopsis thaliana* cytochrome P450-type hydroxylase. The hydroxylase can be used to identify brassinosteroids or brassinosteroid inhibitors, useful to produce plants with modified physiological and/or phenotypic characteristics. The modified plants may show, e.g., stimulation of

CC growth, increased cell elongation, increased wood production,
CC accelerated seed germination at low temperatures, an increase in
CC dry weight, repressed anthocyanin production during growth in light
CC and/or inhibited de-etiolation which is induced, e.g. by cytokinin,
CC in the dark or an increase in stress tolerance. The hydroxylase or
CC its coding sequence can also be used for the recombinant production
CC of compounds, e.g. teasterone.

Sequence 472 AA;

Query Match	21.6%;	Score 552.5;	DB 18;	Length 472;
Best Local Similarity	29.9%;	Pred. No. 3e-42;		
Matches 145; Conservative	98;	Mismatches 205;	Indels 37;	Gaps 14;

[illegible]

RESULT	9
AA044571	
ID	AA044571 standard; Protein; 472 AA

AC AAG44571;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55847.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

Arabidopsis thaliana.

PN EP1033405-A2

AA
PD 06-SEP-2000

25-FEB-2000: 2000EP-0301439

XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 21.6%; Score 552.5; DB 21; length 472;
Best Local Similarity 29.9%; Pred. No. 3e-42;
Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

QY 19 LLEFLAVKLMWEMIMRRVDPNCRSPPLPGMGLPIGETIQLI-----LQRRKFLMKRQ 73
Db 7 llllssiaagfllllrrtrr-rmqllppslqlpligetfqligayktenpepfidera 65
QY 74 KYGCIYKTHLFGNPTVRVMGADNVRQILGHEHKLVSQWPAVATILIGSDTILSNHGVQH 133
Db 66 rygsvmthlfgeptlfsadpetnrfvlqnegklfecsyasiscnllgkshlllmkyslh 125
QY 134 KMKKKAIMRAFSDAL--EHYIPVIOQEVKSAIQEWLQKDCVLYVPEMKKLMERIAMRI 191
Db 126 k-rmhsltmsfansslikdhlmdidrlvrlnldsw---ssrvlmeaekkitfcltvkq 181
QY 192 LLGFEP----EQIKTDEQELVEAFEEEMIKNLFSLPIDVPEFSGLYR-GLRARNFHSKTEE 246
Db 182 lmsfdpgweselrkeyllviegf-----fslplpl-fsttyrkaigarrkvaaltv 233
QY 247 NIRKKIQDDNENEQYKDALQLLIENSRSDEPFSLOAMKEAATELLFGHETASTAT 306
Db 234 vvmkr-reeeegaeerkkmlaal-----aaddgfsdeeiavdlvalvayettstlmt 288
QY 307 SLVMFLGLNTEVVOYKREVEQKVEVMGYTPGKGLSMELDLQKTYGCVIKETLRINPV 366
Db 289 lavkfltetplalaqlkee-hekir-amksdsyslewsdyksmpftgcvnnetlrvanii 346
QY 367 PGGERVALKTFELNGYQIPKGMNVYISICDTHADVDFPNKEERQPERFMSKGLIEDSRE 426
Db 347 ggvframtdvelkykipkwkvfsftravhldpnhfkdartfnprwgsnsavtgcpsn 406
QY 427 NYIPEGGSRMCVGEKFAVLLKIFLVELLQHCHNWILSNGPITMKGTPTIYPVDNLPTKF 486
Db 407 vtfpfgggprlcpypelaravalsvflhrlvtgfsvw-----paedklvifpftlrqkry 461
QY 487 TSYVR 491

Db 462 pifvk 466
RESULT 10
ID AAG45022 standard; Protein; 472 AA.
AC AAG45022;
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56469.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 21.68; Score 552.5; DB 21; Length 472;
Best Local Similarity 29.9%; Pred. No. 3e-42;
Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;
QY 19 LFLAALKWEMIMRVDPNCRSPPLPPGTMGLPFIGETLOL-----LÖRKFLRKQ 73
Db 7 llllsiaagflllllrrtyr-rmgllppsglglpligetfqligayktenpepfiderva 65
QY 74 KYGCIYKTHLFGNPTVRVMGADNVRQILLGEHKLVSVQWPASVRTILGSDTLSNHGVQH 133
Db 66 rygsvfmthlfgeptlsadpetnrfrvlqnegkllfecsyaspasicnllqkxslmlmkgsln 125

OY		134	KNKKKAIMRAFSRDAL--EHYIPVIOQEVKSAIQEWLQKDCVLVYPKKLMFRIAMRI	191
Db		126	k-rmshlmsfiansslikdhlmdldrlvrfnldsw---ssrvllmeaakkittfeltvkg	181
OY		192	LIGFEF---EQIKTDEQLVEAFEEEMIKNLFSLPIDVPFSGLYR-GLRARNFHISKIEE	246
Db		182	lmsfdpgwesesikeylllvlegf-----fslplpl-fsttyrkaigarbkvaeealv	233
OY		247	NIRKKIÖDDDNENEOKYKDALÖLLIENSRSRSDPEPFLQAMKEATELLFGHETTASTAT	306
Db		234	vvmkr-reeeegeaerkkdm]aal1---aaddgfsdeeivdlvallvagyetstimt	288
OY		307	SLVMELGLNTEVVOKVREEVOEKVEMGYTPGKL\$MELLDOLKYTGCVIKETLRINPPV	366
Db		289	lavkfilteptlaqlkee-hekir-amksdsyslewsdyksmpftqcvvneltlrvani1	346
OY		367	PGGFRRVALKTFFELNGYOIPKGWNVIVSICDTHDVADVFPNKEEFQPERFMSKGLEDGSRF	426
Db		347	ggvfirramtdveikgykipgwkvfssfravhldpnhfkdartfnpwrqsnsvltgpsn	406
OY		427	NYIPEGGSGSRMCVGKEFAKVLLKIFVELTQHCHNWILSNGPPTMKGTPIYPVDNLPTKF	486
Db		407	vtfepfgggprlpcpyelarvalsvfhlrvltgfsw----paeqdklvffptrtqtqky	461
OY		487	TSYVR 491	
Db		462	plfvk 466	
 RESULT 11 AAG45021 ID AAG45021 standard; Protein; 491 AA.				
XX	AAG45021;			
DT	18-OCT-2000	(first entry)		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 56468.			
XX				
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.			
OS	Arabidopsis thaliana.			
XX				
PN	EP1033405-A2.			
XX				
PD	06-SEP-2000.			
XX				
PF	25-FEB-2000;	2000EP-0301439.		
XX				
PR	25-FEB-1999;	99US-0121825.		
PR	05-MAR-1999;	99US-0123180.		
PR	09-MAR-1999;	99US-0123548.		
PR	23-MAR-1999;	99US-0125788.		
PR	25-MAR-1999;	99US-0126264.		
PR	29-MAR-1999;	99US-0126785.		
PR	01-APR-1999;	99US-0127462.		
PR	06-APR-1999;	99US-0128234.		
PR	08-APR-1999;	99US-0128714.		
PR	16-APR-1999;	99US-0129845.		
PR	19-APR-1999;	99US-0130077.		
PR	21-APR-1999;	99US-0130449.		
PR	23-APR-1999;	99US-0130510.		
PR	23-APR-1999;	99US-0130891.		
PR	28-APR-1999;	99US-0131449.		
PR	30-APR-1999;	99US-0132048.		
PR	30-APR-1999;	99US-0132407.		
PR	04-MAY-1999;	99US-0132484.		
PR	05-MAY-1999;	99US-0132485.		
PR	06-MAY-1999;	99US-0132486.		
PR	06-MAY-1999;	99US-0132487.		
PR	07-MAY-1999;	99US-0132863.		

PR	11-MAY-1999;	99US-0134256;
PR	14-MAY-1999;	99US-0134218;
PR	14-MAY-1999;	99US-0134219;
PR	14-MAY-1999;	99US-0134221;
PR	14-MAY-1999;	99US-0134370;
PR	18-MAY-1999;	99US-0134768;
PR	19-MAY-1999;	99US-0134941;
PR	20-MAY-1999;	99US-0135124;
PR	21-MAY-1999;	99US-0135353;
PR	24-MAY-1999;	99US-0135629;
PR	25-MAY-1999;	99US-0136021;
PR	27-MAY-1999;	99US-0136392;
PR	28-MAY-1999;	99US-0136782;
PR	01-JUN-1999;	99US-0137222;
PR	03-JUN-1999;	99US-0137528;
PR	04-JUN-1999;	99US-0137502;
PR	07-JUN-1999;	99US-0137724;
PR	08-JUN-1999;	99US-0138094;
PR	10-JUN-1999;	99US-0138540;
PR	10-JUN-1999;	99US-0138847;
PR	14-JUN-1999;	99US-0139119;
PR	16-JUN-1999;	99US-0139452;
PR	16-JUN-1999;	99US-0139453;
PR	17-JUN-1999;	99US-0139492;
PR	18-JUN-1999;	99US-0139454;
PR	18-JUN-1999;	99US-0139455;
PR	18-JUN-1999;	99US-0139456;
PR	18-JUN-1999;	99US-0139457;
PR	18-JUN-1999;	99US-0139458;
PR	18-JUN-1999;	99US-0139459;
PR	18-JUN-1999;	99US-0139460;
PR	18-JUN-1999;	99US-0139461;
PR	18-JUN-1999;	99US-0139462;
PR	18-JUN-1999;	99US-0139463;
PR	18-JUN-1999;	99US-0139750;
PR	18-JUN-1999;	99US-0139763;
PR	21-JUN-1999;	99US-0139817;
PR	22-JUN-1999;	99US-0139899;
PR	23-JUN-1999;	99US-0140354;
PR	23-JUN-1999;	99US-0140353;
PR	24-JUN-1999;	99US-0140695;
PR	28-JUN-1999;	99US-0140823;
PR	29-JUN-1999;	99US-0140991;
PR	30-JUN-1999;	99US-0141287;
PR	01-JUL-1999;	99US-0141842;
PR	01-JUL-1999;	99US-0142154;
PR	02-JUL-1999;	99US-0142055;
PR	06-JUL-1999;	99US-0142390;
PR	08-JUL-1999;	99US-0142603;
PR	09-JUL-1999;	99US-0142920;
PR	12-JUL-1999;	99US-0143297;
PR	13-JUL-1999;	99US-0143542;
PR	14-JUL-1999;	99US-0143624;
PR	15-JUL-1999;	99US-0144005;
PR	16-JUL-1999;	99US-0144085;
PR	16-JUL-1999;	99US-0144086;
PR	19-JUL-1999;	99US-0144325;
PR	19-JUL-1999;	99US-0144331;
PR	19-JUL-1999;	99US-0144332;
PR	19-JUL-1999;	99US-0144333;
PR	19-JUL-1999;	99US-0144334;
PR	19-JUL-1999;	99US-0144335;
PR	20-JUL-1999;	99US-0144332;
PR	20-JUL-1999;	99US-0144632;
PR	20-JUL-1999;	99US-0144884;
PR	21-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0145086;
PR	22-JUL-1999;	99US-0145085;
PR	22-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145089;
PR	23-JUL-1999;	99US-0145192;
PR	23-JUL-1999;	99US-0145195;

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 21.6%; Score 552.5; DB 21; Length 491;
Best local Similarity 29.9%; Pred. No. 3.2e-42;
Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

QY 19 LLEFLAAVKLWEMIMIRVDPNCRSPPLPPGTMLPFIGETLQLI-----LQRRKFLMKRKQ 73
Db 26 LLLLSIAAGFLILLIRTRYR-rmgllppsglpllgelfqligayktenpepfiderva 84
QY 74 KYGCIYKTHLFGNPTVRVMGADNVRQILLGEHKLVSQVMPASVRTILGSDTLSNVHGVOH 133
Db 85 rygsvfmtlhfgeptlfsadpetnrfvlgneqklfecsyasicnllgkshllmkyslh 144
QY 134 KNKKKAIMRAFSPRAL--EHYIPVIOQEVKSAIOEWLQKDCVLYVPEMKKIMFRIAMRI 191
Db 145 k-rmhsllmsfansiikdhlmldidrlvrfnldsw---srvlimeakkitteltvkq 200
QY 192 LLGFEP---EQIKTDEQELVEAFEEEMIKNLSLPIDVPFSGLYR-GLRARNFISKIEE 246
Db 201 lmsfdpgwsestlkeylliviegf-----fslplpl-fsttyrkaigarkeyaaltv 252
QY 247 NIRKKIODDDNENOKYKDALQLLIENSRSDPEFSLQAMKEATELLFGGHETASTAT 306
Db 253 vvmkr-reeeegaerkkmlaall----aaddgfsdeivdlvalilvagyetstlmt 307
QY 307 SLVMFLGLNTEVYQKREEVQEKVEKMGWTPPGKGLSMELLDQIKYTGCVIKETLRNPV 366
Db 308 lavkfltetplaqikee-hekir-amksdsyslewsdykempftqcvnetlrvani 365
QY 367 PGGFVALKTFELNGYQIPKGNVITYSICDTHDVADEVFPNKEEQPERFMSKGLDEGSRF 426
Db 366 ggvframtaveikgykikpgwkvfsfravhldpnhfkdartfnprwgsnsvtqpsn 425
QY 427 NYIPEGGSRMCVGEFAKVLKIFLVELLQHCHNWLISNGPPTMTGPTIYPVDNLPTKF 486
Db 426 vftpfgygprlcpyelaravalsvflhrivtgfsww---paeqdklvffltretqkry 480
QY 487 TSYVR 491
Db 481 plfvk 485
RESULT 12
AAG44570
ID AAG44570 standard; Protein; 492 AA.
XX AAG44570;
AC AAG44570;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55846.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 21.6%; Score 552.5; DB 21; Length 492;

Best Local Similarity 29.9%; Pred. No. 3.2e-42;
Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

QY 19 LLEFLAVKLMWEMLMIRVDPCRSPLPPGTGMLPFIGETLQLI-----LQRRKFLRMKRQ 73
Db 27 lllsslaagflllrtrtyr-rmgllpgslgllrlgetftllgayktenpepfiderva 85
QY 74 KYGCIYKTHLFGNPTVRMGADNVRQILLGEHKLVSQVMPASVRTTIGSDTLNVHGVQH 133
Db 86 rygsvfmthlfgptfsadpetnrflvlgneqklfecsyasicnllgkshllmkysl 145
QY 134 KKKKAIMRAFSRDAL--EHYIPVIOEVKSAIQEWLQKDSCLVYPPEMKKLMFRIAMRI 191
Db 146 k-rmhsltmsfianssllkdhlmldidrlvrfnldsw---ssrvllmeaakkitteltvkq 201
QY 192 LIGFEP---EQIKTDEQLVEAFEEEMIKNLFSLPIDVPFSGLYR-GLRARNFIHSKIEE 246
Db 202 lmsfdpgweselslrkeyllviegf-----fslplpl-fsttyrkaigarckvaaltv 253
QY 247 NTRKKIQDDNENEQKXDALQLLIENSRRSDPEFSIQAMKEAATELLFGHETTASTAT 306
Db 254 vvmkr-reeeegeaekkdmlaall---aaddgfsdeeiavdlvalvagyettstlmt 308
QY 307 SLVMEFLGINTEVQKVRVEQKVMEMGYTPGKGLSMELLQDKYTGCVIKETLRINPV 366
Db 309 lavkfltetplajqlkee-hekir-amksdsyslewsdyksmpftgcvvnetlrvani 366
QY 367 PGCFRVALKTFELNGYQIPKGMNVISICDPHDVADVFPNKKEFPQPERFMSKLEDGSRF 426
Db 367 ggytframtveikgykjpkwkfvssfravhldpnhfkdartfnprwqsnsvttgpsn 426
QY 427 NYIPFGGSRMCVGEKFAKVLKIFLVELTQHCHNWLISNGPPTMKTGPTIYVVDNLPTKF 486
Db 427 vftpfpggprlcpgyelarvalsvflhrlvvgfsw-----paegdklvffpttrtkry 481

QY 487 TSYVR 491
Db 482 pifvk 486

RESULT 13
AAG44572
ID AAG44572 standard; Protein; 444 AA.
XX AC AAG44572;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 55848.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KM termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.

PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141187.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142803.
PR	08-JUL-1999;	99US-0142920.
PR	09-JUL-1999;	99US-0142977.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144816.
PR	21-JUL-1999;	99US-0145086.
PR	22-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147208.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.

PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	30-AUG-1999;	99US-0151080.
PR	31-AUG-1999;	99US-0151303.
PR	01-SEP-1999;	99US-0151438.
PR	07-SEP-1999;	99US-0151930.
PR	10-SEP-1999;	99US-0152363.
PR	13-SEP-1999;	99US-0153070.
PR	15-SEP-1999;	99US-0153718.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154479.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161993.
PR	28-OCT-1999;	99US-0161992.
PR	29-OCT-1999;	99US-0162142.

Query Match	20.9%;	Score 536;	DB 21;	Length 444;
Best Local Similarity	30.0%;	Pred. No. 9.1e-41;		
Matches 138; Conservative	92;	Mismatches 194;	Indels 36;	Gaps 13;

```
QY      44 LPPGTMGLPFGETLÖLI-----LÖRRKFLRMKRÖKGYCITYKTHLEGNPTVRVMGADNVR 98
      ||||:||||| ||| ||| : :|| : |||| ||| : |
Db      3 lppgsllgpligetfqligayktenpepfidervarysvfmthlfgseptlfsadpetnr 62
      ||||:||||: ||| :|| : || : :|| : :|| : |
QY      99 ÖLLGEHKIVSVÖWPASVRTILGSDTLNVHGÖHKNNKKAIMRAFSRDAL--EHYIPVI 156
      :| | || :|||: :|| :|| : :|| : :|| : |
```


Db 63 fvlqneqklfecsyasicnllgkhslllmkgsllhk-rmhsllmsfannsllkdhlmldl 121
QY 157 QOEKSAIQEWLQKSCVLYVPEMKKLMRIALILGFEP----EQIKTDEQELVEAFE 212
Db 122 dlrvfnldsw---ssrvllmeaakkltfcltvkqlmsfdpgwseslrkeylllviegf- 177
QY 213 EMIKNLFSLPIDVPEFSGLYR-GLRARNFHSKIEENIRKKIQDDNENEQKKYKDALQLLI 271
Db 178 -----fslplpl-fstlyrkaigarikaaalvavmkr-reeeegaerkkdmlaal 229
QY 272 ENSRSDPEFSLQAKAEATELLFGHETASTATSLVMLGLNTEVYQKREEVQEKVE 331
Db 230 ----aaddgfsdeelvdlvalivaagyetstclmtlavkfltetplalaqlkee-hektr 284
QY 332 MGMTTPGKGLSMELLDQLKTYGCVIKETLRINPPVPGGFRAVAKTPELNGYQIPKGMNVI 391
Db 285 -amksdsyslswsdgksmpftgcvvnetlrvanliygvrframtveikgkikpgkwvf 343
QY 392 YSICDTHDVADVPFNKEEFQPERFMSKGLEDSRFNYPFGGGSRCVCKEFAKVLKIF 451
Db 344 ssfravhldpnhfkdaftfnprwgsnsvtgtgsnvftpfggprlcpgyelarvalsvf 403
QY 452 LVELTQHCHWILSNCPPTMTGPTIYPVDNLPFTSTYVR 491
Db 404 lhrlvgtgsw----paegdklvfptlttqtkrypifvk 438

RESULT 14

AAG45023
ID AAG45023 standard; Protein; 444 AA.

AC AAG45023;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 56470.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR	26-JUL-1999;	99US-01455276
PR	27-JUL-1999;	99US-01455913
PR	27-JUL-1999;	99US-01455918
PR	27-JUL-1999;	99US-01455919
PR	28-JUL-1999;	99US-01455951
PR	02-AUG-1999;	99US-01463866
PR	02-AUG-1999;	99US-01463889
PR	02-AUG-1999;	99US-01463889
PR	03-AUG-1999;	99US-0147038
PR	04-AUG-1999;	99US-01473204
PR	04-AUG-1999;	99US-0147302
PR	05-AUG-1999;	99US-0147192
PR	05-AUG-1999;	99US-0147260
PR	06-AUG-1999;	99US-0147303
PR	06-AUG-1999;	99US-0147416
PR	09-AUG-1999;	99US-0147493
PR	09-AUG-1999;	99US-0147935
PR	10-AUG-1999;	99US-0148317
PR	11-AUG-1999;	99US-0148319
PR	12-AUG-1999;	99US-0148341
PR	13-AUG-1999;	99US-0148565
PR	13-AUG-1999;	99US-0148684
PR	16-AUG-1999;	99US-0149368
PR	17-AUG-1999;	99US-0149375
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	20-AUG-1999;	99US-0149723
PR	20-AUG-1999;	99US-0149929
PR	23-AUG-1999;	99US-0149902
PR	23-AUG-1999;	99US-0149930
PR	25-AUG-1999;	99US-0150565
PR	26-AUG-1999;	99US-0150884
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154039
PR	20-SEP-1999;	99US-0154779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156458
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	08-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0159293
PR	13-OCT-1999;	99US-0159294
PR	13-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	18-OCT-1999;	99US-0159584
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match	20.9%;	Score 536;	DB 21;	Length 444;
Best Local Similarity	30.0%;	Pred. No. 9.1e-41;		
Matches 138; Conservative	92;	Mismatches 194;	Indels 36;	Gaps 13;

```

QY 44 LPPGTMLPFGTETLÖLI-----LÖRRKFLRMKRÖKYGCICYKTHLLFGNPTVRVMGADNVR 98
      |||:||||| ||| ||| : : : ||| ||| : |
      lppsglpligetfqligayktenpepfidervarysvfmthl fgeptifsaqpetnr 62
Db 3

QY 99 ÖILLGEHKLVSVQMPASVRTILGSDTLNVHGÖHKNNKKAIIMRAFSDAL--EHYIPVI 156
      :| ||| :||| :||| : ||| : : : ||| : |
      63 fvlqnegkifecsypasicnllgkhslllmkyslhk -rmhsilmsfansaslikdhmlidi 121
Db

QY 157 QÖEVKSAIQEWLÖKDSCVLVYPPEMKLMFRIAMRILGFEP---EQIKTDEÖELVEAFE 212
      :| : : | ||| : ||| : : : ||| : ||| : |||
      122 dlrvrfnldsvr---ssrvllmeeakktfelvtvkqlmsfdgsweslrkeyllvlegf- 177
Db

QY 213 EMIKNLFSLPIDVPFESGLYR-GLRARNFHISKIEENIRKKIÖDDNENEÖQKYKDALÖLLI 271
      ||||: || ||| :||| : : : ||| : ||| : |||
      178 -----fstlprpl-fstlyrkaigaarrrkvaeaaltvvyvmkr-reeeeagaerkkdmalaall 229
Db

QY 272 ENSRSDPEPSIQAMKEATELLFGGHETTASTATSLVMFLGINTEVVÖKVREEOEKEVE 331
      :||: ||| : : : ||| : ||| : ||| : ||| : |||
      230 ----aaddgfsdeeivdflvallvagyettsttimtlavkfletetplalaqlkee-hekir 284
Db

QY 332 MGYTPGKGLSMELLDÖLKYTGCYVAKETLRINPPVPGGFRAALKTEFELNGYÖIPKGMNVI 391
      | : : : : : ||| : ||| : ||| : ||| : ||| : |||
      285 -amksdsyslewsdykssmpftgcvvynetlrvanliggvfirramtdveikykipgkwkvf 343
Db

QY 392 YSICDTHVADVFPNNKEEFÖPEREMSKGLEDSRFNYIPREGGSRMCVQGEFAKYLIIKF 451
      | | : : : : ||| : ||| : ||| : ||| : ||| : |||
      344 ssfravhldpnhfkdartfnpwrwgsnsvltgpsnvftfpgggprlcpgyelaraalsvf 403
Db

QY 452 LVELTÖHCNMWLLSNGPPTMKTGPTIYPPVDNLPTKFTSYVR 491
      | | : : : : ||| : ||| : ||| : ||| : |||
      404 lhrvlvtgfsww-----paegdklvfiptrtrtqkrypilfvk 438

```

RESULT	15
ID	AAB07921
AC	AAB07921 standard; Protein; 513 AA.
XX	
XX	AAB07921;
DT	14-NOV-2000 (first entry)
XX	
DE	A cytochrome P450 enzyme designated DWF4.
XX	
KW	DWF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
XX	plant phenotype; cell elongation.
OS	Arabidopsis sp.
XX	
PN	W0200047715-A2.
XX	
PD	17-AUG-2000.
XX	
PF	11-FEB-2000; 2000WO-US03820.
XX	
PR	11-FEB-1999; 99US-0119657.
XX	11-FEB-1999; 99US-0119658.
PA	(ARIZ-) ARIZONA BOARD OF REGENTS.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 10:41:47 ; Search time 88.14 Seconds
(without alignments)
125.614 Million cell updates/sec

Title: US-09-668-482-2
Perfect score: 2563
Sequence: 1 MGLYTLMTVFLCTIVLPVLL.....GPTIYPVDNLPKFTSYVRN 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	2563	100.0	492	3 US-08-724-466B-2	Sequence 2, Appli
2	2563	100.0	492	4 US-08-882-164D-2	Sequence 2, Appli
3	1738	67.8	497	3 US-08-724-466B-4	Sequence 4, Appli
4	1738	67.8	497	4 US-08-882-164D-4	Sequence 4, Appli
5	1734	67.7	497	4 US-08-882-164D-32	Sequence 32, Appli
6	552.5	21.6	472	2 US-08-622-166A-2	Sequence 2, Appli
7	552.5	21.6	472	2 US-08-622-166A-4	Sequence 4, Appli
8	343	13.4	511	4 US-08-991-677-4	Sequence 4, Appli
9	339.5	13.2	490	1 US-08-201-118-3	Sequence 3, Appli
10	339.5	13.2	490	1 US-08-201-118-9	Sequence 9, Appli
11	339.5	13.2	490	2 US-08-238-821B-3	Sequence 3, Appli
12	339.5	13.2	490	2 US-08-238-821B-9	Sequence 9, Appli
13	339.5	13.2	490	5 PCT-US95-05744-3	Sequence 3, Appli
14	339.5	13.2	490	5 PCT-US95-05744-9	Sequence 9, Appli
15	330.5	12.9	520	2 US-09-091-432-2	Sequence 2, Appli
16	322.5	12.6	500	2 US-08-314-601-2	Sequence 2, Appli
17	322.5	12.6	500	5 PCT-US95-13051-2	Sequence 2, Appli
18	316.5	12.3	490	1 US-08-201-118-13	Sequence 13, Appli
19	316.5	12.3	490	2 US-08-238-821B-13	Sequence 13, Appli
20	316.5	12.3	490	5 PCT-US95-05744-13	Sequence 13, Appli
21	311.5	12.2	490	1 US-08-201-118-1	Sequence 1, Appli
22	311.5	12.2	490	2 US-08-238-821B-1	Sequence 1, Appli
23	311.5	12.2	490	5 PCT-US95-05744-1	Sequence 1, Appli
24	302	11.8	510	3 US-08-948-564-4	Sequence 4, Appli
25	296	11.5	496	1 US-08-313-075A-50	Sequence 50, Appli
26	293.5	11.5	490	1 US-08-201-118-7	Sequence 7, Appli
27	293.5	11.5	490	2 US-08-238-821B-7	Sequence 7, Appli

28	293.5	11.5	490	5 PCT-US95-05744-7	Sequence 7, Appli
29	287	11.2	504	1 US-08-457-274A-25	Sequence 25, Appli
30	287	11.2	504	5 PCT-US95-05758-25	Sequence 25, Appli
31	282.5	11.0	476	1 US-08-313-075A-30	Sequence 30, Appli
32	275.5	10.7	500	4 US-09-292-768-68	Sequence 68, Appli
33	274.5	10.7	500	4 US-09-292-768-4	Sequence 4, Appli
34	271.5	10.6	500	3 US-08-881-784-9	Sequence 9, Appli
35	271.5	10.6	500	4 US-09-292-768-70	Sequence 70, Appli
36	268.5	10.5	490	1 US-08-201-118-11	Sequence 11, Appli
37	268.5	10.5	490	2 US-08-238-821B-11	Sequence 11, Appli
38	268.5	10.5	490	5 PCT-US95-05744-11	Sequence 11, Appli
39	268.5	10.5	516	3 US-08-948-564-12	Sequence 12, Appli
40	268.5	10.5	521	3 US-08-948-564-14	Sequence 14, Appli
41	267.5	10.4	490	1 US-08-201-118-5	Sequence 5, Appli
42	267.5	10.4	490	2 US-08-238-821B-5	Sequence 5, Appli
43	267.5	10.4	490	5 PCT-US95-05744-5	Sequence 5, Appli
44	267.5	10.4	496	4 US-09-292-768-64	Sequence 64, Appli
45	266	10.4	513	3 US-08-948-564-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-724-466B-2
Sequence 2, Application us/08724466B
Patent No. 6063606
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
ZIP: M5L 1A9
COUNTRY: Canada
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-2

Query Match 100.0%; Score 2563; DB 3; Length 492;
Best Local Similarity 100.0%; Pred. No. 1.8e-246;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLYTLMTVFLCTIVLPVLLFLAAVKLMWLMIRVDNCRSPLPPTMGILPFIGETIQL 60
DB 1 MGLYTLMTVFLCTIVLPVLLFLAAVKLMWLMIRVDNCRSPLPPTMGILPFIGETIQL 60

QY	61	ILORRKFRLMRKROKYYGCIYKTHLFGNPTVYRWGADNVROILLGEHKLVSQWPA5VRTIL	120
Db	61	ILQRRKFLMRKROKYGCITYKTHLFGNPTVYRWGADNVROILLGEHKLVSQWPA5VRTIL	120
QY	121	GSDTL5NVHGVOHKNNKKKAIMRAFSRDALEHYIPVIOQEVKSAIOEWLOKDSCLVYYPEM	180
Db	121	GSDTL5NVHGVOHKNNKKKAIMRAFSRDALEHYIPVIOQEVKSAIOEWLOKDSCLVYYPEM	180
QY	181	KLIMERIAMRILLGFEPEQIKTDEQELVEAFEEIMIKNLFSLPIDVPFSGLYRGLARNEFI	240
Db	181	KLIMERIAMRILLGFEPEQIKTDEQELVEAFEEIMIKNLFSLPIDVPFSGLYRGLARNEFI	240
QY	241	HSKIEENIRKKIQQDDNENENQKKYDALQULLIENSRSDPEPFSLOAMKEATELLEFGGHET	300
Db	241	HSKIEENIRKKIQQDDNENENQKKYDALQULLIENSRSDPEPFSLOAMKEATELLEFGGHET	300
QY	301	TASTATSLVMFLGLNTEVVOQKAVREVOEKVEMGMYTPGKGLSMELLDQLKYTGCVIKETL	360
Db	301	TASTATSLVMFLGLNTEVVOQKAVREVOEKVEMGMYTPGKGLSMELLDQLKYTGCVIKETL	360
QY	361	RINPVPVGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADVFPNKKEEQPERFMSKGL	420
Db	361	RINPVPVGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADVFPNKKEEQPERFMSKGL	420
QY	421	EDGSRFNYIPFGGGSRCMVGRKEFAKVLKIFLVELLQHCNMWIL5NGPPTMKTGPTIYPVD	480
Db	421	EDGSRFNYIPFGGGSRCMVGRKEFAKVLKIFLVELLQHCNMWIL5NGPPTMKTGPTIYPVD	480
QY	481	NLPTKFTSYVRN 492	
Db	481	NLPTKFTSYVRN 492	

```

RESULT      2
US-08-882-164D-2
: Sequence 2, Application US/08882164D
: Patent No. 6306624
:
: GENERAL INFORMATION:
:   APPLICANT:  Petkovich, P. Martin, White, Jay A.,
:   APPLICANT:  Beckett, Barbara R., Jones, Glenville
:   TITLE OF INVENTION:  Retinoid Metabolizing Protein
:   NUMBER OF SEQUENCES:  43
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE:  Blake, Cassels & Graydon
:     STREET:     Box 25, Commerce Court West
:     CITY:       Toronto
:     STATE:      Ontario
:     COUNTRY:    Canada
:     ZIP:        M5L 1A9
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE:  Diskette, 3 1/2 inch, 1.4 Mb storage
:   COMPUTER:     COMPAQ, IBM PC compatible
:   OPERATING SYSTEM:  MS-DOS 5.1
:   SOFTWARE:     WORD PERFECT
:
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER:  US/08/882,164D
:   FILING DATE:       June 25, 1997
:
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER:  08/667,546
:     FILING DATE:       June 21, 1996
:     APPLICATION NUMBER:  08/724,466
:     FILING DATE:       October 1, 1996
:
:   ATTORNEY/AGENT INFORMATION:
:     NAME:  Hunt, John C.
:
:   REGISTRATION NUMBER:  36,424
:   REFERENCE/DOCKET NUMBER:  50767/00010
:   TELECOMMUNICATION INFORMATION:
:     TELEPHONE:  (416) 863-4344
:     TELEFAX:   (416) 863-2653
:
:   INFORMATION FOR SEQ ID NO:  2:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH:  492 amino acids
:       TYPE:    amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-882-164D-2

```

Query Match	100.0%;	Score 2563;	DB 4;	Length 492;
Best Local Similarity	100.0%;	Pred. No. 1.8e-246;		
Matches 492;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MGLYTLMVTEFLCTIVLPVLLFLAAYKIMEMLMIRVDPNCRSPBPGTMGLPIGETIQL	60
Dd	1	MGLYTIMVTFLCTIVLPVLLFLAAVKIMELMIRVDPNCRSPBPBTGMGLPIGETIQL	60
QY	61	ILQRKKFLRMKRÖKYGCYYKTHLFGNPTRVMGADNVRÖILLGEHKLVSVÖWPASVRTIL	120
Dd	61	ILQRKKFLRMKRÖKYGCYYKTHLFGNPTRVMGADNVRÖILLGEHKLVSVÖWPASVRTIL	120
QY	121	GSDTLSNVHGVOHKNNKKAIMRAFSRDALHHYIPVIÖQEYKSAIÖEWLÖKDSCVLYPEM	180
Dd	121	GSDTLSNVHGVOHKNNKKAIMRAFSRDALHHYIPVIÖQEYKSAIÖEWLÖKDSCVLYPEM	180
QY	181	KKLMEFIAMRILLGFEBEQIKTDEQELVEAFEEMIKNLFSLPIDVPFSGLYRGRLARNF I	240
Dd	181	KKLMEFIAMRILLGFEBEQIKTDEQELVEAFEEMIKNLFSLPIDVPFSGLYRGRLARNF I	240
QY	241	HSKIEENIRKKIÖDDDNENEÖKYKDALÖLLIENSRRSDEPESLÖAMKEATELLEFGHET	300
Dd	241	HSKIEENIRKKIÖDDDNENEÖKYKDALÖLLIENSRRSDEPESLÖAMKEATELLEFGHET	300
QY	301	TASTATSLVMFLGNTENVVÖKVAREVÖEKEVEMGMTPGKGLSMELLDÖLKTYTCVIAKETL	360
Dd	301	TASTATSLVMFLGNTENVVÖKVAREVÖEKEVEMGMTPGKGLSMELLDÖLKTYTCVIAKETL	360
QY	361	RINPPVPGGFRVALKTFFELNGYÖLPKGWNVYISICDTHADVFPNKKEFÖPEREMSGL	420
Dd	361	RINPPVPGGFRVALKTFFELNGYÖLPKGWNVYISICDTHADVFPNKKEFÖPEREMSGL	420
QY	421	EDGSRENIYPGGGSRMCVGKEFAKVLKIFVELTELÖHCNWILSNGPTMKTGPTIYPVD	480
Dd	421	EDGSRENIYPGGGSRMCVGKEFAKVLKIFVELTELÖHCNWILSNGPTMKTGPTIYPVD	480
QY	481	NLPPTKFTSYVRN 492	
Dd	481	NLPPTKFTSYVRN 492	

RESULT 3
 US-08-724-466B-4
 ; Sequence 4, Application US/08724466B
 ; Patent No. 6063606
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: Petkovich, P. Martin, White, Jay A.,
 ; APPLICANT: Beckett, Barbara R., Jones, Glenville
 ; TITLE OF INVENTION: Retinoid Metabolizing Protein
 ; NUMBER OF SEQUENCES: 30
 ;
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Blake, Cassels & Graydon
 ; STREET: Box 25, Commerce Court West
 ; CITY: Toronto
 ; ZIP: M5L 1A9
 ;
 ; COUNTRY: Canada
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
 ; COMPUTER: COMPAQ, IBM PC compatible
 ; OPERATING SYSTEM: MS-DOS 5.1
 ; SOFTWARE: WORD PERFECT
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/724,466B
 ; FILING DATE: October 1, 1996
 ;
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/667,546
 ; FILING DATE: June 21, 1996
 ;
 ; ATTORNEY/AGENT INFORMATION:
 ;

NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-4

Query Match 67.8%; Score 1738; DB 3; Length 497;
Best Local Similarity 68.0%; Pred. No. 2.2e-164;
Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;

QY 1 MGLYTLMTFLCTIVLPVLLFLAAVKLMMLIRVDPNCRSPPLPGTMGLPFIGETLQL 60
Db 1 MGLPALLASALCTFVLPPLFLAIAIKLMDLYCVSGRDRSCALPLPGTMGFPFGETLQM 60
QY 61 ILQRRKFLMRKQKQYGCITYKTHLFGNPTVRVMGADNVKQILLGEHKLVSVQWPASVRTIL 120
Db 61 VLQRRKFLQMKRRKRYGYIKYKTHLFGRPVVRVMGADNVKRIILGDDRLVSVHWPASVRTIL 120
QY 121 GSDTLNVHGVQHKNNKKAIMRAFSRDALHHYIPVIOQEVKSAIOEWLQKDSK---VLV 176
Db 121 GSGCLSNLHDSHKQKKVIMRAFSREALCYVPVITEEVGSSLEQWL---SCGERGLLV 177
QY 177 YPEMKLMFRIAMRILLGFEPEQIKTD---EOELVEAFEEEMIKNLFSLPIDVPFSGLYRG 233
Db 178 YPEVKRLMFIAMRILLGCEP-QLAGDGSEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRG 236
QY 234 LRARNFIHSKIEENIRKKIQD-DDNENEOKKDALQLLIENSRRSDEPFSLQAMKEATE 292
Db 237 MKARNLIHARIEQNIIRAKICGLRASEAGQCKDALQLLIHSWGERLDMQALKOSTE 296
QY 293 LIFGHEHTASTATSLVMFLGNTENVQKVEEVQEKVEMGYTPGKGLSMELLDQKYT 352
Db 297 LIFGHEHTASTATSLITYLGLYPHVLQKVRBELKSKGLLCKSNQDNKLDMEILEQLKYI 356
QY 353 GCVIKETLRINPVPVGGFRVALKTFELNGYQIPKGNVYISICDTHDVADVPFNKEEFQ 412
Db 357 GCVIKETLRINPVPVGGFRVALKTFELNGYQIPKGNVYISICDTHDVAEIFTNKEEFNP 416
QY 413 ERFMSKGLDEGSRFNYPFGGSRMCVGEFAKVLKIFLVELTQHCNWLISNGPPTMKT 472
Db 417 DRFSAPHPEASRFSFIPFGGLRSCVGEFAKILKIFVELARHCDWQOLLNGPPTMKT 476
QY 473 GPTIYPVDNLPTKFTSY 489
Db 477 SPTVYYPVDNLPARFTHF 493

RESULT 4
US-08-882-164D-4
; Sequence 4, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESS: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-4

Query Match 67.8%; Score 1738; DB 4; Length 497;
Best Local Similarity 68.0%; Pred. No. 2.2e-164;
Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;

QY 1 MGLYTLMTFLCTIVLPVLLFLAAVKLMMLIRVDPNCRSPPLPGTMGLPFIGETLQL 60
Db 1 MGLPALLASALCTFVLPPLFLAIAIKLMDLYCVSGRDRSCALPLPGTMGFPFGETLQM 60
QY 61 ILQRRKFLMRKQKQYGCITYKTHLFGNPTVRVMGADNVKQILLGEHKLVSVQWPASVRTIL 120
Db 61 VLQRRKFLQMKRRKRYGYIKYKTHLFGRPVVRVMGADNVKRIILGDDRLVSVHWPASVRTIL 120
QY 121 GSDTLNVHGVQHKNNKKAIMRAFSRDALHHYIPVIOQEVKSAIOEWLQKDSK---VLV 176
Db 121 GSGCLSNLHDSHKQKKVIMRAFSREALCYVPVITEEVGSSLEQWL---SCGERGLLV 177
QY 177 YPEMKLMFRIAMRILLGFEPEQIKTD---EOELVEAFEEEMIKNLFSLPIDVPFSGLYRG 233
Db 178 YPEVKRLMFIAMRILLGCEP-QLAGDGSEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRG 236
QY 234 LRARNFIHSKIEENIRKKIQD-DDNENEOKKDALQLLIENSRRSDEPFSLQAMKEATE 292
Db 237 MKARNLIHARIEQNIIRAKICGLRASEAGQCKDALQLLIHSWGERLDMQALKOSTE 296
QY 293 LIFGHEHTASTATSLVMFLGNTENVQKVEEVQEKVEMGYTPGKGLSMELLDQKYT 352
Db 297 LIFGHEHTASTATSLITYLGLYPHVLQKVRBELKSKGLLCKSNQDNKLDMEILEQLKYI 356
QY 353 GCVIKETLRINPVPVGGFRVALKTFELNGYQIPKGNVYISICDTHDVADVPFNKEEFQ 412
Db 357 GCVIKETLRINPVPVGGFRVALKTFELNGYQIPKGNVYISICDTHDVAEIFTNKEEFNP 416
QY 413 ERFMSKGLDEGSRFNYPFGGSRMCVGEFAKVLKIFLVELTQHCNWLISNGPPTMKT 472
Db 417 DRFSAPHPEASRFSFIPFGGLRSCVGEFAKILKIFVELARHCDWQOLLNGPPTMKT 476
QY 473 GPTIYPVDNLPTKFTSY 489
Db 477 SPTVYYPVDNLPARFTHF 493

RESULT 5
US-08-882-164D-32
; Sequence 32, Application US/08882164D
; Patent No. 6306624

GENERAL INFORMATION:

APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-32

Query Match 67.7%; Score 1734; DB 4; Length 497;
Best Local Similarity 67.7%; Pred. No. 5.4e-164;
Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

QY 1 MGLYTMVTEFLCTIVLPVLLFLAAVKIMWMIMIRVDNCRSPPLPPTGMLPFIETLQL 60
DB 1 MGLPALIASALCTFVLLPLFLAALKLMDLYCVSSRDRSCALPLPPTMGFPFEGETLQM 60
QY 61 ILQRRKFLRMKROKYGCIYKTHLFGNPTVRVMGADNVROILLGEHKLVSQWPASVRTL 120
DB 61 VLQRRKFLQMKRRKYGFTYKTHLFGRTVRVMGADNVRRILLGEHRLVSVMHPASVRTL 120
QY 121 GSDTLNVHGVQHKRKKKAIMRAFSRDALENHYIPVIOQEVKSAIQEWLQKDSCLLV 176
DB 121 GAGCLSNLHDSHQRKKVIMQAFSREALQCYVLIAEEVSSCLEQWLSCGERGLLV 177
QY 177 YPEMKKIMFRIAMRILLGFEPEQI-KTDEQELVEAFEEEMIKNLFSLPIDVFPFSGLYRGL 234
DB 178 YPEVKRIMFRIAMRILLGCEPAGGEGEQQLVFAFEEMTRNLFSLPIDVFPFSGLYRGV 237
QY 235 RARNFIHSKIEENIRKKIQD-DDNENEQYKALQOLLIEHSRSDSEPFSLQAMKEATEL 293
DB 238 KARNLIHARIEENIRAKIRRLQATEPDDGGCDALQOLLIEHSWGERLDMQALKQSTEL 297
QY 294 LFGHETTASTATSLVMFLGLNTEVVOQVREVEQEKVEMGYTPGKGLSMELLDQLKYTG 353
DB 298 LFGHETTASATSLITYLGLYPHVLQKYREBKIKSGLLCKSNQDNKLDMEETLEQLKYIG 357
QY 354 CVIKETLRINPVPGGFRVALKTFELNGYQIPKGMNVISICDTHDVADVPFNKEEFOPE 413
DB 358 CVIKETLRINPVPGGFRVALKTFELNGYQIPKGMNVISICDTHDVADIFTNKEEFNPD 417

QY 414 REMSKGLEDSRENTYIPFGGSRMCVGEKFAVLLKIFLVELLQHCNMTISNGPPTMTKG 473
DB 418 RFIVPHPEDASRPSFIPFGGLRSCVGEKFAKILLKIFTVELARHCDWOLLNGPPTMTKS 477
QY 474 PTIYPVDNLPTKFTSY 489
DB 478 PTVYPVDNLPARTTYE 493

RESULT 6

US-08-622-166A-2
Sequence 2, Application US/08622166A
Patent No. 5952545
GENERAL INFORMATION:
APPLICANT: KONCZ, CSABA
APPLICANT: MATHUR, JAIDEEP
APPLICANT: SZEKERES, MIKLOS
APPLICANT: ALTMANN, THOMAS
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,166A
FILING DATE: 27-MAR-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0147-0153P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-622-166A-2

Query Match 21.6%; Score 552.5; DB 2; Length 472;
Best Local Similarity 29.9%; Pred. No. 1.7e-46;
Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

QY 19 LLFLAAVKLMWEMIMIRVDNCRSPPLPPTGMLPFIETLQLI-----LQRRKFLRMKRO 73
DB 7 LLLLSIAAGFLLLRTRR-RMGLPPGSLGLPIGETFQILGAYKTEBPPIDERVA 65
QY 74 KYGCIYKTHLFGNPTVRVMGADNVROILLGEHKLVSQWPASVRTLGSDDLNVHGVQH 133
DB 66 RYGSVEMTHLFGEPITFSADPETNRFVLQNEGKLFECSSYPASICNLGKHSLLMKGSLH 125
QY 134 KKKKAIMRAFSRDAL--EHYIPVIOQEVKSAIQEWLQKDSCLVVPPEMKKIMFRIAMRI 191
DB 126 K-RMHSLSMFSANSSIIRKDHMLDIDRLVRFNDSW--SSRVLMEEAKKITFELTVKQ 181
QY 192 LLGFEP---EQIKTDEQELVEAFEEEMIKNLFSLPIDVFPFSGLYR-GLRARNFIHSKIEE 246
DB 182 IMSFDPGEWSESLRKEYLLVIEGF-----FSLPLPL-FSTTYRKAIQARRKVAEALTV 233

OY	247	NIRKTIODDNEENÖKYKDALÖLLIENSRRSDEPFSLÖAMKEATELLFGGHETASTAT	306
	:	: : : : : : : : : : : : :	
Dd	234	VVMKR-REEEEGAERKKDMIALI---AADDGFSDEEIVDFVALLVAGYEFTSTIMT	288
OY	307	SLVMFLGLNTEVVÖKVREEVÖQEKYEMGYTPRGKLSMELLDÖLKYTCGVIKETLRINEPV	366
	:	: : : : : : : : :	
Dd	289	LAVKFLETFPLALÖLKEE-HEKIR-AMKSDSYSLEWSDYKSMPTÖCVNETLRVANIT	346
OY	367	PGGFRVALKTFLNGYÖIPKGWNVTYSICDTHDVADVFPRNKEEQPERFMSKLGEDSRF	426
	:	: : : : : : :	
Dd	347	GGVFRRAMTVDEIKGYKIPKGWKVSSFRVHLDPNHFKDARTFNPMRWÖSNSVTTGPSN	406
OY	427	NYIFEGGSRMÖVGKEFAKVLKLIFVELTQHNCMWILNSGRPMTKGTPTYVDNLPTRF	486
	: :	: : : : : : : : :	
Dd	407	VFTPEGGPRLCPGYELARVALSVFLHRLVTGFSGSW-----PAEDDKLVFPPTTRTÖQRX	461
OY	487	TSYVR 491	
	: :		
Dd	462	PIFVK 466	

```

1      RESULT 7
2      US-08-622-166A-4
3      ; Sequence 4, Application US/08622166A
4      ; Patent No. 5952545
5      ;
6      ; GENERAL INFORMATION:
7      ;
8      ; APPLICANT: KONCZ, CSABA
9      ; APPLICANT: MATHUR, JAIDEEP
10     ; APPLICANT: SZEKERES, MIKLOS
11     ; APPLICANT: ALTMANN, THOMAS
12     ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
13     ; TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS I
14     ; NUMBER OF SEQUENCES: 4
15     ; CORRESPONDENCE ADDRESS:
16     ; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
17     ; STREET: P.O. Box 747
18     ; CITY: Falls Church
19     ; STATE: Virginia
20     ; COUNTRY: USA
21     ; ZIP: 22040-0747
22     ; COMPUTER READABLE FORM:
23     ; MEDIUM TYPE: Floppy disk
24     ; COMPUTER: IBM PC compatible
25     ; OPERATING SYSTEM: PC-DOS/MS-DOS
26     ; SOFTWARE: Patentn Release #1.0, Version #1.30 (EPO)
27     ; CURRENT APPLICATION DATA:
28     ; APPLICATION NUMBER: US/08/622,166A
29     ; FILING DATE: 27-MAR-1996
30     ; CLASSIFICATION: 800
31     ; ATTORNEY/AGENT INFORMATION:
32     ; NAME: SVENSSON, LEONARD R.
33     ; REGISTRATION NUMBER: 30,330
34     ; REFERENCE/DOCKET NUMBER: 0147-0153P
35     ; TELECOMMUNICATION INFORMATION:
36     ; TELEPHONE: (703) 205-8000
37     ; TELEFAX: (703) 205-8050
38     ; TELEX: 248345
39     ; INFORMATION FOR SEQ ID NO: 4:
40     ; SEQUENCE CHARACTERISTICS:
41     ; LENGTH: 472 amino acids
42     ; TYPE: amino acid
43     ; TOPOLOGY: linear
44     ; MOLECULE TYPE: protein
45     ;
46     ; US-08-622-166A-4

```

Query Match 21.6%; Score 552.5; DB 2; Length 472;
Best Local Similarity 29.98; Pred. No. 1.7e-46;
Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

QY 19 LLFLAAVKLWEMLMIRVDPNCRSPLPPTGMLPFIGETLLQLI-----LQRKFLRMKRQ 73
|||||:::| | |||||::| | | | | | | : :

[illegible]

```

RESULT      8
US-08-991-677-4
; Sequence 4, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent L
; APPLICANT: Carraway, Daniel T
; APPLICANT: Smeltzer, Richard H
; TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
; FILE REFERENCE: 50617
; CURRENT APPLICATION NUMBER: US/08/991,677A
; CURRENT FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: US 60/033,381
; EARLIER FILING DATE: 1996-12-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Liquidambar styraciflua
US-08-991-677-4

```

[illegible]

QY	185	FRAMRILLGFEPEQIKTDEQELVEAFEE--MIKNLESLPIDVPF-SGLYRGLRAR----	237
		: : : : : : : :	
Db	187	YRAAFGTI-----SHEDQDEFVALIQEFSQLEGAFNIADFIPLKWPQGINVRLNKA	239
QY	238	-----NEIHSKIEENIRKKIQQDDNENEQYKDALQLLIENSRSDEPFSIQAMK-----	287
		::: : : : : :	
Db	240	RGALDGFIDKIIDDHIQKSGKNSSEYVDTMVDLLAFYGEFAKVSESDDLQNSIKLTKN	299
QY	288	-EATELLFGGHETTASTATSLVMFLGNTVEVQKREVEQKVEMGMYTPKGJLSMELL	346
		: ::: : : ::: :	
Db	300	IKAIMDVMFGGTETVASAIEWAMTELMKSPEDLKKVQOELAVVVGGLDRVVEEKD-----F	354
QY	347	DOLKYTGCVIKETLRINPVPVPGGFRAVALKTEFLNGYQIIPKGNVYISICDTHVDVDFPN	406
		:: : ::: :	
Db	355	EKLTYLKCVLKEVLRHLHPPIPLLHETAEDAEVGGYIIPAKSRVMINACAIGRKNSWAD	414
QY	407	KEEFQPERFMSKGLD--GSRFNYPFGGSRMCVCGKEFAKVLKIFLVELTQHC-NWIL	463
		: : : : : :	
Db	415	PDTFRPSRFLKDGVPDFKGNNEFIPIFGSGRRSCPGMQLGLYALETTVAHLH-HCFTWEL	473
QY	464	SNG 466	
		:	
Db	474	PDG 476	

```

RESULT      9
; US-08-201-118-3
; Sequence 3, Application US/08201118
; Patent No. 5786191
; GENERAL INFORMATION:
; APPLICANT: GOLDSTEIN, Joyce A.
; APPLICANT: ROMKES-SPARKS, Marjorie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
; TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
; TITLE OF INVENTION: SUBFAMILY
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/201,118
; FILING DATE: 22-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,962
; FILING DATE: 09-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebeschuetz, Joe
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 15280-192-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 490 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-201-118-3

```

Query Match	13.2%;	Score 339.5;	DB 1;	Length 490;
Best Local Similarity	26.9%;	Pred. No. 2.9e-25;		
Matches 139;	Conservative 83;	Mismatches 218;	Indels 77;	Gaps 22;

```

QY 5 TLMVTEFCTIVLPEVLLFLAAVKLWEMLMIRVDPNCRSPPLPGCMTGLPIGETTQL-ILQ 63
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 SLVVLVLC---LSCILLLS---LWR-----QSSGRGKLPGPPTPLPVIGNILQIGIKD 49

QY 64 RRKFLRMKRQKQYGCITYKTHLFGNPTVRMGADNVROIL-LGEHKLVSVOVPASVRTILG 121
   . | | : : : : : | | | | | | | | | | | | | | | | | | | | | |
Db 50 ISKSLTNLSKVYGPVFTLYFGLKPIVVLHGYEAVKEALIDIGEEFSGRGIFPLAERANRG 109

QY 122 SDTLSNVHGVQHKNNKK-AIMRAFSRDALEHYIPVIOQEVKSAIQEWLOKDSCVLYPEM 180
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110 -----FGIYFSGKKWKEIRFSLMTLRNF----GMGRKSIEDRVQEEARCIPEELR 157

QY 181 KKLMEIRIAMRILGFEPEOI-----KTDEQ--ELVEAFEEMIK-----N 217
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 158 KTKASPCDPFTILGCAPCNVICSIIFHKRFDYKDOQFLNIMEXLNIENIKILSSPWIQICN 217

QY 218 LFSLPIDVPFSGLYRGLRARN--FJHSKIEENIRKKIQQDDNENENQKYKDALQILLIENS 275
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 218 NFSPIIDY-FPGTHNKL-LKNVAEMKSYILEKVEKHQESMDNNPNQDIDCFLMKMEKEK 275

QY 276 RSD-EPEFSLQAMKEATELLFGGHETTASTATSLVMFLGLNTEVVOKVREVEQKEMGM 334
   : : : : : : | : | | | | | | | | | | | | | | | | | | | | | |
Db 276 HNQPSFEFTIESLENTAVDLFGAGFETSTTLRYALLLLKHPYATAKVQEEIERI----- 331

QY 335 YTPGKGLS--MELLDQLKYTGCVATKETLRINPVPVPGGFYALK-TFELNGYQIDPKGWNVI 391
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 332 ---GRNRSPCMQDRSHMPYTDAVVEYQRYLDLLPTSLPHAVTCDIKFRNYLILPKGTTIL 388

QY 392 YSICDT-HDVADVFPNKEEFQEPHFMSKGLLEDGSRFN----YIPFGGSRMCYQKEFAKV 446
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 389 ISLTSVYLHDNKE-FPNPEMFDPHNF----LDEGGENEKSKYFMPFSAGKRIVGEALAGM 443

QY 447 LKIFIVELTQHCCNMWILSNGPPTMKTGPTIYPVDNLP 483
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 444 ELFLFSLTLQNFNLKSLVDPKMLDTPPVVNGFASVP 480

```

RESULT 10
 US-08-201-118-9
 : Sequence 9, Application US/08201118
 : Patent No. 5786191
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: GOLDSTEIN, Joyce A.
 :
 : APPLICANT: ROMKES-SPARKS, Marjorie
 :
 : TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
 :
 : TITLE OF INVENTION: DNAS FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
 :
 : TITLE OF INVENTION: SUBFAMILY
 :
 : NUMBER OF SEQUENCES: 44
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: Townsend and Townsend Kourie and Crew
 :
 : STREET: 379 Lytton Avenue
 :
 : CITY: Palo Alto
 :
 : STATE: California
 :
 : COUNTRY: US
 :
 : ZIP: 94301
 :
 : COMPUTER READABLE FORM:
 :
 : MEDIUM TYPE: Floppy disk
 :
 : COMPUTER: IBM PC compatible
 :
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 :
 : SOFTWARE: Patentln Release #1.0, Version #1.25
 :
 : CURRENT APPLICATION DATA:
 :
 : APPLICATION NUMBER: US/08/201,118
 :
 : FILING DATE: 22-FEB-1994
 :
 : CLASSIFICATION: 435
 :
 : PRIOR APPLICATION DATA:
 :
 : APPLICATION NUMBER: US 07/864,962
 :
 : FILING DATE: 09-APR-1992
 :
 : ATTORNEY/AGENT INFORMATION:
 :
 : NAME: Liebeschuetz, Joe

```

;      REGISTRATION NUMBER: 37,505
;      REFERENCE/DOCKET NUMBER: 15280-192-1
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (415) 326-2400
;      TELEFAX: (415) 326-2422
;      INFORMATION FOR SEQ ID NO: 9:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 490 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;
US-08-201-118-9

```

Query Match	13.2%;	Score 339.5;	DB 1;	Length 490;
Best Local Similarity	26.9%;	Pred. No. 2.9e-25;		
Matches 139; Conservative	83;	Mismatches 218;	Indels 77;	Gaps 22;

[illegible]

RESULT 11
 US-08-238-821B-3
 ; Sequence 3, Application US/08238821B
 ; Patent No. 5912120
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: GOLDSTEIN, Joyce A.
 ; APPLICANT: ROMKES-SPARKS, Marjorie
 ; APPLICANT: DE MORAIS, Sonia M.F.
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
 ; TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
 ; TITLE OF INVENTION: MEPHENTHOIN METABOLISM
 ; NUMBER OF SEQUENCES: 61
 ;
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: California
 ;

```

: COUNTRY: US
: ZIP: 94111
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/238,821B
: FILING DATE: 06-MAY-1994
:
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/201,118
: FILING DATE: 22-FEB-1994
:
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-192110US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 326-2400
: TELEFAX: (650) 326-2422
:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 490 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-08-238-821B-3

```

Query Match	13.2%;	Score 339.5;	DB 2;	Length 490;
Best Local Similarity	26.9%;	Pred. No. 2.9e-25;		
Matches 139;	Conservative 83;	Mismatches 218;	Indels 77;	Gaps 22;

QY	5	TLMTVFCTIVLPVLLFLAAVKIMMLIRVDPNCRSPLPGTMGLPIGELLQI-LIQ	63
		: :	
Db	3	SLVVLVLC---LSCLLLS---LWR-----QSSGRKLPBGPPTLPVIGNIIOIGIKD	49
QY	64	RRKFLRMKROKYGCIYKTHLFGNPTVRVMGADNVROIL--LGEHKLVSVOVPASVFTILG	121
		: : : : : : : : : : : :	
Db	50	ISKSLTNLSKYYGPVFTLYFGKLPYVLHGVEAVKALIDLGEESGRCIGFPLAERANRG	109
QY	122	SDTLSNHVGOHKNNKK-AIMRAFSDALAHYIPVIOEYKSAIOEWLOKDSCLVYPFM	180
		: : : : : : : : : :	
Db	110	-----FGIVSNGKKWKEIRRFSLMTLRNF-----GMGRSIEDRVQEEARCIVEELR	157
QY	181	KKLMFRIAMRILGFEPEQI-----KTDEQ--ELVEAFEEMIK-----N	217
		: : : : : : : : : : :	
Db	158	KTKASPCDDPTFLGCAPCNVICSIIFHKRFYKDKQOFLNMEKLNENIKILSSPWIOICN	217
QY	218	LESPLIDVPESGLYRGLRARN--FIHSKIEENIRKKIQDDNENEQKYKDALOLLIENS	275
		: : : : : : : : :	
Db	218	NFSPIIDY-FPGTHNKL-LKNVAFMKSTYLEKVEKEHOESMDMNNPODFIDCFMKMEKEK	275
QY	276	RSD-EPFSLQAMKEATELLFGGHETTASTATSLVMFLGLNTEVYQVREVEQKEVEMG	334
		: : : : : : : : : : : : :	
Db	276	HNQPSFTIESLENTAVDLFGAGTETTTSTLRALLLLKHPEVYAKVQEEIERVI----	331
QY	335	YTPGKGLS--MELLDQLKYTGCVIKETLRINBPVPGGERVALK-'TFELNGYQIPKGMNVI	391
		: : : : : : : : : : :	
Db	332	---GRNRSPCMODRSHMPYTDVAVHVORYLDLLPTSLPHAVTCDIKFRNYLIPKGTITL	388
QY	392	YSICDT-HDVADVFPNKEEFQPERFMSKGLDGSRFN---YIPGGGSRMCVQKEFAKY	446
		: : : : : : : : :	
Db	389	ISLTSVLHDNKE-FPNEMFDPNHF---LDEGGNEFKSKYFMFESAGKRICVGEALAGM	443
QY	447	LKIFLVELTQHCHNWILSNGRPMTKTGPTIYVVDNLP	483

Db 444 ELFELTSLQNFNLKSLVDPKNDLTPVNGFASVP 480

RESULT 12
US-08-238-821B-9

; Sequence 9, Application US/08238821B
; Patent No. 5912120

; GENERAL INFORMATION:

; APPLICANT: GOLDSTEIN, Joyce A.

; APPLICANT: ROMES-SPARKS, Marjorie

; APPLICANT: DE MORAIS, Sonia M.F.

; TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN

; TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-

; TITLE OF INVENTION: MEPHENTHIN METABOLISM

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: US

; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/238,821B

; FILING DATE: 06-MAY-1994

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 07/864,962

; FILING DATE: 09-APR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Liebeschuetz, Joe

; REGISTRATION NUMBER: 37,505

; REFERENCE/DOCKET NUMBER: 15280-192110US

; TELEPHONE: (650) 326-2400

; TELEFAX: (650) 326-2422

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 490 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-238-821B-9

Query Match 13.2%; Score 339.5; DB 2; Length 490;

Best Local Similarity 26.9%; Pred. No. 2.9e-25;

Matches 139; Conservative 83; Mismatches 218; Indels 77; Gaps 22;

QY 5 TLMVTFCTIVLPVLFLLAALKMELMIRKVPNCRSPPLPGTGLPFIGETQL-ILQ 63

Db 3 SLVVLVYC---LSCLLLS---LMR-----QSSGRKLPFGPTPLPVIGNITLQIGIKD 49

QY 64 RRKFLRMKRKOKYGCITKTHLEGNPTVVRMGADNVKQIL--LGEHKLVSVOVPASVRIILG 121

Db 50 ISKSLTNLSKVGYPVFTLYFGKLPVVLHGYEAVKKEALIDGEEFSGRGIFPLAERANRG 109

QY 122 SDTLSNVHGVQHKKKK-AIMRAFSRDALAHYIPVIOQEVKSAIQEWLQKDSCLVYYPEM 180

Db 110 -----FGIVFSNGKKWKEIRPSIMTLRNF-----GMGKRISIEDRVQEEARCLVEELR 157

QY 181 KKLMEIRIAMRILGFEPQI-----KTDEQ--ELVEAFEEEMIK-----N 217

Db 158 KTKASPCDPTFILGCAPCNVICSIIFHKRPDYKDQQLNLMKINENIKILSSPWIOJCN 217

QY 218 LFSLPIDVPFSGLYRGLRARN--FTHSKIEENIRKKIQQDDNENQKYKDALQLLIENSR 275

Db 218 NFSPIIDY-FPGTHNKL-LKNVAFMKSYILLEKVEKHQESMDMNNPQDFIDCFLMKMEREK 275

QY 276 RSD-EPFSLQKKEATELLFGGHETASTATSLVMFLGLNTEVQKRVREVEQKEVMGM 334

Db 276 HNQPSFTIESLNTAVDLFGAGTETSTLRVALLLLKHPEVTAKVQEEIERVI----- 331

QY 335 YTPGKGLS--MELLQKLTGCVIKETLRINPVPVPGGFVALK-TFELNGYQIPKGNVYI 391

Db 332 ---GRNRSPCMQDRSHMPTDAVVEVQRYIDLPTSLPHAVICDIFRNYLIPKGTIL 388

QY 392 YSICDT-HDVADYFPNKEEFQPERFMSKLEDGSRFN---YIPFGGSRMCVQKEFAKY 446

Db 389 ISLTVLHDNKE-FPNPEMFDPHNF----LDEGNGFKSKYFMPFSAGKRICVGEALAGM 443

QY 447 LKIFLVELTQHONWILSNPPTMKGTPTIYPVDNLP 483

Db 444 ELFELTSLQNFNLKSLVDPKNDLTPVNGFASVP 480

RESULT 13
PCT-US95-05744-3

; Sequence 3, Application PC/TUS9505744

; GENERAL INFORMATION:

; APPLICANT: GOLDSTEIN, Joyce A.

; APPLICANT: ROMES-SPARKS, Marjorie

; APPLICANT: DE MORAIS, Sonia M.F.

; TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN

; TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT

; TITLE OF INVENTION: OF S-MEPHENTHIN METABOLISM

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend Kourie and Crew

; STREET: 379 Lytton Avenue

; CITY: Palo Alto

; STATE: California

; COUNTRY: US

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/05744

; FILING DATE:

; CLASSIFICATION:

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/238,821

; FILING DATE: 06-MAY-1994

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/201,118

; FILING DATE: 22-FEB-1994

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 07/864,962

; FILING DATE: 09-APR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Dow, Karen B.

; REGISTRATION NUMBER: 29,684

; REFERENCE/DOCKET NUMBER: 15280-192-1-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 326-2400

; TELEFAX: (415) 326-2422

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 490 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

;; CURRENT APPLICATION NUMBER: US/09/091,432
;; CURRENT FILING DATE: 1998-06-18
;; EARLIER APPLICATION NUMBER: PCT/US96/20094
;; EARLIER FILING DATE: 1996-12-19
;; EARLIER APPLICATION NUMBER: US 60/009,119
;; EARLIER FILING DATE: 1995-12-22
;; EARLIER APPLICATION NUMBER: US 60/013,388
;; EARLIER FILING DATE: 1996-03-14
;; NUMBER OF SEQ ID NOS: 3
;; SOFTWARE: Microsoft Word 2.0C
;; SEQ ID NO 2
;; LENGTH: 520
;; TYPE: PRT
;; ORGANISM: Artificial
;; FEATURE:
;; OTHER INFORMATION: Sequence is deduced from DNA sequence of SEQ ID NO:1
;; Patent No. 5981837
US-09-091-432-2

Query Match 12.9%; Score 330.5; DB 2; Length 520;
Best Local Similarity 24.9%; Pred. No. 2.5e-24;
Matches 125; Conservative 89; Mismatches 207; Indels 81; Gaps 20;

QY 13 TIVLPVLLFLAAVKLMEMIRRVDPNCRSPPLPGTGLPFGTLLQLLQRRKFLMRK 72
DB 18 SLVIVVSLF-----IFSFTRRRRPYPGPRGWPPIGNMIMMDQLTHRGIANLA 68
QY 73 OKYG--CIYK---THLFGNPTVRVAGADNVRQILLGHNKLVSVQWPAASVR---TILGSD 123
DB 69 KKYGGLCHLRMGFLHMYA-----VSSPEVARQVLQVDSVFSNR-PATIAISYLTIDRAD 122
QY 124 TLSNVHG-VQHKNNKKAIMRAFSRDALEHYIPVIOEVKSAIQEWLQKDSK----VLVY 177
DB 123 MAFAHYGPFWQMRKVCVMKVSRRKRAESW----ASVRDEVDMVRSVSCNVGKPIVNG 177
QY 178 PEMKKLMFRIAMRILLGFEPEQIKTDQELVEAFEEEMIK--NLSPLIDVPFSGLY---- 231
DB 178 EQIFALTFRNITYRAAFSACEK--GQDEFIRILOEFSKLFGAFNVADEFIPYFGWIDPQG 234
QY 232 ---RGLRARN---FIHSKIEENIRKK-----IODDNEENEQKYKDALQLLIENSRRSDE 279
DB 235 INKRLVKARNDLDFIDDIIDEHMKKENQNAVDDGDVVDVDDLLAFYSEAKLVSE 294
QY 280 PFSLO-----AMKEATELLFGHETASTATSLVMFLGLNTEVVOKVREVOEKVE 331
DB 295 TADLQNSIKLTRDNITKAIIMDVMEGGTETVASAIEWALTTELRSPEDLKRVOELAEVY- 353
QY 332 GMYTTPGKLSMEL---LDQLKYTGCVIKETLRINPPVPGGFRVALKTFELNGYQIPKG 387
DB 354 -----GLDRRVEESDIEKLTFLKCTLKEITLRMHPPIPLLHETAEDTSIDGFFIPKK 405
QY 388 WNVYISICDTHDVADVFPNKEEFQPERFMSKLED--GSRFNYIPFGGSRMCVGEFAK 445
DB 406 SRVMINAFALGRDPTSWTDPDTFRPSRPLEPGVPDFKGSNFEFTIPFGSGRRSCPGMQLGL 465
QY 446 VLLKIFLVELTQHC-NWILSNG 466
DB 466 YALDLAVAHIL-HCFTWKLPDG 486

Search completed: November 6, 2001, 13:27:04
Job time: 9917 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:25:32 ; Search time 71.14 Seconds
(without alignments)
526.819 Million cell updates/sec

Title: US-09-668-482-2
Perfect score: 2563
Sequence: 1 MGLYTLMTFLCTIVLPVLL.....GPTIYPVDNLPFKFTSYVRN 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	689.5	26.9	444	1	S75761	cytochrome P450 -
2	552.5	21.6	472	1	S55379	cytochrome P450 CY
3	532.5	20.8	485	2	A84859	probable cytochrom
4	503	19.6	482	2	T02739	probable cytochrom
5	503	19.6	513	2	T46143	steroid 22-alpha-h
6	489.5	19.1	457	2	T04444	cytochrome P450 -
7	481	18.8	464	2	T07859	cytochrome P450 ho
8	455	17.8	457	2	D85429	cytochrome P450 li
9	450	17.6	490	2	H86185	hypothetical prote
10	394.5	15.4	489	2	B84733	probable cytochrom
11	382.5	14.9	512	2	H96759	probable steroid 2
12	375	14.6	519	1	T02263	cytochrome P450 DW
13	363	14.2	492	1	D70649	cytochrome P450 RV
14	356.5	13.9	495	2	T06288	probable cytochrom
15	349	13.6	490	2	T06286	probable cytochrom
16	347.5	13.6	516	2	T48140	flavonoid 3',5'-hy
17	343.5	13.4	501	2	A34236	cytochrome P450 3A
18	339.5	13.2	490	2	B38462	S-mephenytoin 4-hy
19	339.5	13.2	490	2	D28951	cytochrome P450 2C
20	339.5	13.2	501	2	A29487	cytochrome P450 3A
21	338	13.2	503	2	JC4702	cytochrome P450 3A
22	335	13.1	497	2	H84712	probable cytochrom
23	334	13.0	520	2	T48141	flavonoid 3',5'-hy
24	330.5	12.9	520	2	T04591	ferulate-5-hydroxy
25	329	12.8	504	2	S50892	cytochrome P450 3A
26	328.5	12.8	530	1	A27491	lanosterol 14alpha
27	326	12.7	503	2	JC7243	sterol 14alpha-dem
28	326	12.7	511	1	A34260	laurate omega-hydr
29	326	12.7	526	2	T04566	cytochrome P450 ho

30	324.5	12.7	504	2	A29410	cytochrome P450, 9
31	323.5	12.6	508	1	A36304	cytochrome P450 4A
32	322.5	12.6	500	2	A26685	cytochrome P450 2C
33	322.5	12.6	503	2	S14275	steroid 6beta-mono
34	322	12.6	518	2	T48143	flavonoid 3',5'-hy
35	320	12.5	513	2	JC7120	cytochrome P450 en
36	319.5	12.5	500	2	S62785	cytochrome P450 2C
37	318	12.4	503	2	JC4759	sterol 14-demethyl
38	317.5	12.4	496	2	T52255	cytochrome P450 li
39	317	12.4	486	2	JC4240	lanosterol 14alpha
40	317	12.4	503	2	JC4758	lanosterol 14alpha
41	316.5	12.3	455	2	T48973	cytochrome P450-11
42	316.5	12.3	490	2	F38462	S-mephenytoin 4'-h
43	316	12.3	349	2	A86329	hypothetical prote
44	315	12.3	499	2	T07113	probable cytochrom
45	315	12.3	509	1	S68855	lanosterol 14alpha

ALIGNMENTS

RESULT 1

S75761 cytochrome P450 - Synechocystis sp. (strain PCC 6803)

N:Alternate names: protein slr0574

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: S75761

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S75761

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-444 <KAN>

A:Cross-references: EMBL:D64003; GB:AB001339; NID:g1001200; PIDN:BA010496.1; PID:g100

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: cyp

C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase

F:251-413/Domain: cytochrome P450 homology <P45>

F:391/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 26.9%; Score 689.5; DB 1; Length 444;
Best Local Similarity 34.5%; Pred. No. 1.3e-38;
Matches 156; Conservative 99; Mismatches 170; Indels 27; Gaps 10;

QY	43	P L P G T M G L P F I G E T L Q L I L O R R K F L R M K R O K Y G C I Y K T H L F G N P T V R V M G A D N V R Q I L L	102
DB	12	P I P P G D F G L P W L G E T L N - F L N D G D F G K K R Q Q G F P I F K T R L F G K N V I F I S G A L A N R F L F T	70
QY	103	G E H K L V S V Q W P A S V R T I L G S D T L S N H G V O H K N K K A I M R A F S R D A L E H Y I P V I Q E V K S	162
DB	71	K E Q E T F O A T W P L S T R I L L G P N A L A T O M G E I H R S R K I L Y Q A F L P R I L D S Y L P K M D G I V Q G	130
QY	163	A I Q E W L Q K D S C V L Y P E M K K I M F R I A M R I L L G E P E Q I K T D E Q L V A F E E M I K N L F S L P	222
DB	131	Y L E Q W - G K A N E V I W P Q L R M T F D V A A T L F M G - - - E K V S Q N P Q - L F P W F E T Y I Q G L F S L P	185
QY	223	I D V P E S G L Y R G L R A R N F I S K I E E N I R K K I O D D N E N E Q K Y K D A L O L I E N S R S R D E P F S	282
DB	186	I P L P N T L I F G K S Q R A R A L L A E L E K I I K A R Q Q P S E - - - - E D A L G I L L A A R D N N Q P L S	240
QY	283	L Q A M K E A T E L L F G G H E T T A S T A T S L V M F L G I N T E V V Q K V R E V Q E K V E M G M Y T P G K G L S	342
DB	241	L P E L K D Q I L L L F A G H E T L T S A L S S F C L L L G H S D I R E R V R Q E - Q N K I Q L - - - - - S Q E L T	294

N:Alternate names: cytochrome P450 homolog T914.17
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02739; D84692
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; Ron
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.
A:Reference number: 214710
A:Accession: T02739
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-482 <R0U>
A:Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461849
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: D84692
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: GB:AE002093; NID:g3461849; PID:AAC33235.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29090; T914.17
A:Map position: 2
A:Introns: 80/2; 187/3; 238/3; 321/3; 351/3; 413/3; 453/2
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
F;291-453/Domain: cytochrome P450 homology <P45>

Query Match 19.6%; Score 503; DB 2; Length 482;
Best Local Similarity 28.9%; Pred. No. 4.1e-26;
Matches 140; Conservative 92; Mismatches 224; Indels 28; Gaps 11;

QY 7 MYTFLCTIVLPVLLFLAAVKLWEMLMIRVDPNCRSPPLPGTMGLPFIGETLQILQIR-R 65
Db 22 LITLTIVVVVLLF-----KWLHMKQRL-----RLPPSGMGLPYIGETLRLYTENPN 71
QY 66 KFLRMKROKYGCITYKTHLFGNPTVRVMGADNVRLQILGHNKLVSVQWPASVPTILGSDTL 125
Db 72 SFPATRONKYGDIFKTHILGCPVMISSPEARMYLVSKAHLFKPTYPPEKRMIGPEAL 131
QY 126 SNVHGVOHKKKKAIMRAFSRDALHNYIPVIOQEVKSAIQEWL-QKDSCVLYPEMKKLM 184
Db 132 FTHQGPYHSTLKLRLVQSSFMPSALRPTVSHIELLVQLTSSWTSQKSINTLEY--MKRYA 189
QY 185 FRIAMRILGFEPEDQIKTDEQLVEAFEEEMIKNLSLPIDVPFSGLYRGLRARNFIHSKI 244
Db 190 FDVAIMSAFGDKEEPTTIDYIKLL--YQRLERGYNSMPLDLLPGLFHKSMKAR-----IEL 243
QY 245 EENIRKKIQQDDNENEQKYKDALQLLIENSRSDPEPSIQAMKEATELLFGHETAST 304
Db 244 SEELRKVI-EKRRENGREGLLVLLGAKDQKRNGLSDSOIADNIIGVIFAADTTASV 302
QY 305 ATSLVMEGLNTEVVOQKRVREYQEKVEMGYTPPKGLSMELLQDLKYTGCVIKETLRINP 364
Db 303 LFWLLKYLHDHPNLLQEVSR-EQESIRQIKKENRRISWEDTRKMPLTTRVIOETLRAAS 361
QY 365 PVPGGFVALKTFELNGYQIPKGMNVYISICDTHVDADVFPNKKEEFQPEREMSKLEDGS 424
Db 362 VLSFTFRAVQDVEYDGLIPKMKVLPFRRIHHSSEFPDPEKFDPSRF-----EYAP 416
QY 425 R-FNYIPFGGSGRMCGKEFAKVLKTFVLVELTQHCNWLISNGPPTMTKTPTIYPVDNLP 483
Db 417 KPTYMPFGNGVHSCPGSELAKLEMLILLHLTTTSFRWEVIGDEEGIOYGPFPVPKGLP 476

RESULT 5
T46143
N:Alternate names: cytochrome P450 homolog T914.17
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 16-Feb-2001
C:Accession: T46143
R:Bloeker, H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.; Quetier, F.; S
submitted to the Protein Sequence Database, December 1999
A:Reference number: 223024
A:Accession: T46143
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-513 <BLO>
A:Cross-references: EMBL:AL132979
A:Experimental source: cultivar Columbia; BAC clone T3A5
C:Genetics:
A:Map position: 3
A:Introns: 74/2; 182/3; 233/3; 338/3; 369/3; 396/1; 432/3
A:Note: T3A5.40
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F;308-484/Domain: cytochrome P450 homology <P45>
F;462/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 19.6%; Score 503; DB 2; Length 513;
Best Local Similarity 27.6%; Pred. No. 4.5e-26;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

QY 4 YTLMYTFLCTIVLPVLLFLAAVKLWEMLMIRVDPNCRSPPLPGTMGLPFIGETLQIL-- 61
Db 7 HTLLPLLLPSLLSLFL-----ILKRRNRKTRFNLPPGKSGWPFLETIGYLPK 58
QY 62 ---LQRRKFLRMKROKYGCITYKTHLFGNPTVRVMGADNVRLQILGHNKLVSVQWPASVPT 118
Db 59 YTATTLGDFEMQOHVSKYKTYRSNLFGEPITVSADAGLNRFILQNEGRLEFECSTPSTIGG 118
QY 119 ILGSDTLSNVHGVOHKKKKAIMRAFSRDALH-----YIPVIOQEVKSAIQEWLQKDS 172
Db 119 ILGKWSMLVVGDMHRD-----MRSLNLSLFLSHARLRTILKDYERHTLFLVDSW-QQNS 172
QY 173 CVLVPEMKKIMFRIAMRILGFEPEDQIKTDEQLVEAFEEEMIKNLSLPIDVPFSGLYR 232
Db 173 IFSQODEAKKFTFNLMAKHIMSMDPEEET--EQLKKEYVTFMKGVASAPLNLPGTAYHK 230
QY 233 GLRAR---NFIHSKIEENIRKKIQQDDNENEQKYKDALQLL-----IENSRSD- 279
Db 231 ALQSRATILKFERKMEERKLDIKEEDQEEVEVKTDEAEEMSKSDHVRKQRTDDDLLGVY 290
QY 280 ---PFSLQAMKEATELLFGHETTASTATSLVMEFLGLNTEVVOQKREE---VOEKVE 331
Db 291 LKHSNLSTEQILDLILSLFAGHETTSVAIALAIFPLQACPKAVEELREHLEIARAKKE 350
QY 332 MGMTTPGKGLSMELLQDLKYTGCVIKETLRINPVPVPGGFVALKTFELNGYQIPKGMNVY 391
Db 351 LG---ESELNWDYKKMDFTQCIVINETRLGWNVRFLLHRKALKDVRYKGYDIPSGWKVL 406
QY 392 YSICDTHVDADVFPNKKEEFQPERFM-----SKLEDGSRF--NYIPFGGSGRMCGVKE 442
Db 407 PVISAVHLDNSRYDQPNLENFWRWQQNNGASSSGSFSFTWGNMYMPFGGPRLCAGSE 466
QY 443 FAKVLLKIFLVELTQHCNWLISNGPPTMTKTPTIYPVDNLP 483
Db 467 LAKLEMAVFIIHVLKFNWELAEDDK-----PFAFPVDFP 502

RESULT 6
T04444
cytochrome P450 - Arabidopsis thaliana
N:Alternate names: protein T18B16.200; protein T5K18.10
N:Contains: oxidoreductase (EC 1.-.-.-.-)

C;Species: *Arabidopsis thaliana* (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2000
C;Accession: T04444; T05806
R;Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoerge, W.; Bancroft, I.; Mewes, H.W.
submitted to the Protein Sequence Database, April 1998
A;Reference number: Z15359
A;Accession: T04444
A;Molecule type: DNA
A;Residues: 1-457 <BEV>
A;Cross-references: EMBL:AL021687
A;Experimental source: cultivar Columbia; BAC clone T18B16
R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voelt, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, April 1998
A;Reference number: Z15453
A;Accession: T05806
A;Molecule type: DNA
A;Residues: 131-457 <BEW>
A;Cross-references: EMBL:AL022580
A;Experimental source: cultivar Columbia; BAC clone T5K18
C;Genetics:
A;Map position: 4
A;Introns: 67/2; 173/3; 302/3; 358/1; 393/3
A;Note: T18B16.200; T5K18.10
C;Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase
F;272-433/Domain: cytochrome P450 homology <P45>

Query Match	19.1%	Score 489.5	DB 2	Length 457
Best Local Similarity	28.2%	Pred. No. 3e-25		
Matches 127	Conservative 97	Mismatches 183	Indels 43	Gaps 12
QY	20	LFLLAAVKLM--EMLMIRRVDPNCRSPLPPTGMLPFIGETLLQILLQ--RKFLRMKRQKYG 76		
		: : : : : : : : : : : : : :		
Db	10	LEAGSLFLYFLRCLISQRRFGSSKLPLPPGTGMWPYGETFQYLSQDPNVFFQSKQRYG 69		
QY	77	CLYKTHLFGNPFTVRVMGADNVRQILLGEHKLVSQWPAVSRTILGSDTLNSNVHGQHKNK 136		
		:: : : : : : : : : : : : : : : : : : : : :		
Db	70	SVFKTHVLGCPQVMISSPEAKRFLVTKSHLFKPTFPASKERMLGQALFFHQGDYHAKL 129		
QY	137	KKAIMRAFSRDALAHYIPVIOEYKSAIQEWLQKDSCLVLYPEMKKLMERIAMRILLGFE 196		
		: : : : : : : : : : : : : : : : : : :		
Db	130	RRLVLRAFMPESIRNMVPDIESTAQDSLRSW--EGTMINTYQEKKTYTFENVALLSTFG-- 185		
QY	197	PEQIKTDE---QELVEAFEEMTKNLFSLPIDVPFSGLYRGLARNFTHSKIEENIRKKI 252		
		: : : : : : : : : : : : : : : : : : : : : : :		
Db	186	-----KDEVLYREDLKRCYVILEKGYNSMPVNLPGTLFHKSKAR-----KELSQILLARI 235		
QY	253	QDDNENENEQKYKDAL-QLLIENSRRSDEPFSLOAMKEATELLFGGHETTASTATSLVMF 311		
		: : : : : :		
Db	236	LSERRONGSSHNDLIGSFMGDKELLTDEQIA-----DNLIGVIFAARDTTASVMSWILKY 290		
QY	312	LGINTFVVQKRVREY---QEKVEMGMTTPGKGLSMELLDQLYKTCVIKETLRINPPVP 367		
		: : : : : : : : : : : : : : : : : : : : : :		
Db	291	LAENPNVLEAVTEEQMAIRKDKEE-----GESLTWGDTKKMPLTSRVIOETLLRVASITLS 344		
QY	368	GGFRVALKTFELNGYQIPKGMNVYISICDTHDVADYFPNKEEFQPERFMSKGLDGSRFN 427		
		: : : : : : : : : : : : :		
Db	345	FTFREAVEDEVVEGYLIPKGWKVLPFLFRNIHISADIFSNPKGKDPDSRF-----EVAAPKN 399		
QY	428	-YIPFGGSRMCVCGKEFAVKVLLKIFLVELT 456		
		: : : : : : : : : : : : : : : : : : : :		
Db	400	TFMPFGNGTHSCPGNELAKLEMSIMIHILT 429		

RESULT 7
T07859
cytochrome p450 homolog - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 16-Feb-2001
C:Accession: T07859
R:Bishop, G.J.; Harrison, K.; Jones, J.D.
Plant Cell 8, 959-969, 1996

A;Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes th
A;Reference number: Z16181; MUID:96266705
A;Accession: T07859
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-464 <BIS>
A;Cross-references: EMBL:U54770; NID:g1421740; PIDN:AAB17070.1; PID:g1421741
A;Experimental source: strain GCR758
C;Genetics:
A;Gene: dwarf
C;Superfamily: Synecocystis cytochrome P450 slr0574; cytochrome P450 homology
F;273-436/Domain: cytochrome P450 homology <P45>

	Query Match	18.8%;	Score 481;	DB 2;	Length 464;	
	Best Local Similarity	28.4%;	Pred. No. 1.le-24;			
	Matches 133;	Conservative 90;	Mismatches 192;	Indels 54;	Gaps 13;	
QY	20 LFLAAVKIMWMLMIRVDPNCRSPLEPGTGLPFIGETLQLLLQRKKFLRMKRQKYGCIIY	79	:	:	:	:
Dp	16 IFTCALLRNQVKYNOKN-----LPFGTMGWPLEGETTEFLKLCPSFMKNQARARYGSFF	69	:	:	:	:
QY	80 KTHLFNGPTVRVMGADNVRRQILLGEHLVSVQWPASVRTILGSDTLNSVHGVOHKNNKKA	139	:	:	:	:
Dp	70 KSHILGPTIVSMDSELNRYLVLNEAKGLVPGPQSMIDILGKCNAAVNSAHHKMYRGA	129	:	:	:	:
QY	140 IMRAFS----RDALFHYIPVIQEVKSALQEWLQDKSCVLYVPENKKLMERIAMRIILGF	195	:	:	:	:
Dp	130 LLSTLSPTMIRDQL--LPKIDEFRSHLLTNWDNK--VIDIQENTNMKAFSLSKQIAGI	184	:	:	:	:
QY	196 EPEQIKTDEQELVEAFEEMIKNLFSLPIDVPFSGLYRGLRAR---NFHSKIEE-NIRK	250	:	:	:	:
Dp	185 ESTSL---AQEFMSEEFNLVLTSLPINLPNTNYHRGFQARKTIYNLLRTLIEERRASK	241	:	:	:	:
QY	251 KIQDD-----DNENEQKKDALOLLINSRSRSDPEFSLQAMKEATELLFGGHETTASTA	305	:	:	:	:
Dp	242 EIQHDMGLYLMNEEATRFK-----LTDE-----MDLLITILYSGETYVSTS	285	:	:	:	:
QY	306 TSLVMEFLGLNTEVVQKVREE---VQEKVEMGMTTPGKGLSMELLDQKYTGCVIKETLRI	362	:	:	:	:
Dp	286 MMAVKYLLHDHPKYLEELRKHEMAIREKK-----PEDPIDINDYRSMRFTTRAVILLET SRL	340	:	:	:	:
QY	363 NPYPVGGERVALKTFFELNGYOIPKGWNVIYSICDTHDVADVFPNKKEEQPERFMSKLED	422	:	:	:	:
Dp	341 ATIVNGVLKRTQTQDMEINGIITPKGWRIYVTRELNIYDRLLPYDPYSFNPRWMDSLEH	400	:	:	:	:
QY	423 GSRFNYYIPFGGSRMCVCKEFAKVLLIKIFVELLTQHCNWLLSNGPPTMK	471	:	:	:	:
Dp	401 QNSF--LVFGGGTQCPCGELGVAEISTFLHYFVTKYRWEEIGGDKLWK	447	:	:	:	:

RESULT 8
D85429
cytochrome P450 like protein [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Apr-2001
C/Accession: D85429
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory
Nature 402, 769-777, 1999
A/Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A/Reference number: A85001; MUID:20083488
A/Accession: D85429
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-457 <STO>
A/Cross-references: GB:NC_001268; NID:g7270586; PIDN:CAB80304.1; GSPDB:GN00140
C/Genetics:
A/Gene: AT4g36380
A/Map position: 4
C/Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C/Keywords: heme; iron; metalloprotein
F:396/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	17.8%;	Score 455;	DB 2;	Length 457;
Best Local Similarity	28.6%;	Pred. No. 6.1e-23;		
Matches 130;	Conservative 90;	Mismatches 185;	Indels 50;	Gaps 16;

```

QY 44 LPRGTMLPFIETLQJLIL-----QRRKFLMKRQKYGYCIYKTHLFGNPLVYKMGADNVR 98
   :|:|:| | | | | | | :|:|:| | | | | | | :|:|:| | | | | | | :|:|:| | | | | | |
Db 2 IPNGSLGWPVIGETLNFIAAGYSSRPVTFMDKRKSLYGVKTEKNIIGFPIIISTDAE-VN 60

QY 99 QJLGEHKLVSV-QMPASVRTILGSDTLSNVHGVOHKNNKKAIMRAFSRDALHYIPVIO 157
   :|:| | | :|:| | :|:| | :|:| | :|:| | :|:| | :|:| | :|:| | :|:| | :|:| |
Db 61 KVLVLNHNFTFVPAFKSITELLGENSILSING-PHOKRLHTLIGAFIIRS--PHLKDRIT 117

QY 158 QEVKSAT----QEWLOKDCSCLVYPPMKKLMFRIAMRILLGFEPEQIKTDEQELVEAFEE 213
   :|:|:|:| :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 118 RDIASVLTLASWAOL-PLVHVODEIKKMTFELLVKVLMSTSPGE--DMNILLKEFEE 173

QY 214 MIKNLFLPIDVPFSGLYRGLRARNFHSKIENIRKKIÖDD--NENEQYKDALQJLLI 271
   || | | | | :|:| | | | :|:| | :|:| | :|:| | :|:| | :|:| | :|:| | :|:| |
Db 174 FIKGLICIPIKFPGTRILYKSLKAK---ERLIKVKVKEERQVAMTTSPANDVVDVL 229

QY 272 E---NSRRSDEPFSLQAMKEATELLEFGHETTASTATSLVMFLGLNTEVVOKVREEVÖE 328
   :|:|:| | | :| | | | :|:| | | | | | | | | | | | | | | | | | | |
Db 230 RDGDSSEKÖSPDEVSCK--IVEMMIPGEETMPTAMTIAVKFLSDNPALAKL---VEE 284

QY 329 KVEMGMVTPGKGLSMELLÖJLKYTG-----VIKETLRINPVPVGGFVALKTFELN 380
   :| | :| | :|:| | | | | | | | | | | | | | | | | | | | | | | |
Db 285 NMEM-----KRRKLEGEELYKMTDYMSLSTQNVINETLRMANIINGVWRKALKDVEIK 338

QY 381 GYQIPKGMNVYISICDTHDVADVPNKKEEFQPEREMSKGLJEDGSRFNYIPFGGSRMCVG 440
   || | | | | | | :| | | | :|:| | :|:| | :|:| | :|:| | :|:| | :|:| |
Db 339 GYLIPKGCVLASFISVHDEDDIYDNPYÖFDPWRMDRINGSANSISICFTPFEGGGRICPG 398

QY 441 KEPAKVLLKIFVLVELTQHCHNWILSNGP---PTMK 471
   | :|:| :| | | | :| | | | | | | | | | | | | | | | | | | |
Db 399 LEISKLEISIFLHLVTRYSWTAEEDEIVSFPTVK 433

```

RESULT 9
 H86185
 hypothetical protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: H86185
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: H86185
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-490 <STO>
 A:Cross-references: GB:AE005172; NID:g2388581; PIDN:AAB71462.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match	17.68;	Score 450;	DB 2;	Length 490;
Best Local Similarity	27.68;	Pred. No. 1.4e-22;		
Matches 125; Conservative	84;	Mismatches 220;	Indels 24;	Gaps 10;

```

QY      41 RSP L P G T M G L P F I G E T L Q L I L Q R R ----- K F L R M K R O K Y G -- C I Y K T H L E G N P T V R Y M G 93
      |  | | | | : | | | | | : : : | | | | | : | | | | : :
Db      42 R H Y L P P G D L G W F P I G N M L S F L R A F K T S D P D S F T R T L I K R Y G P K C I Y K A H M F G N P S I I Y T T 101

```

[illegible]

RESULT 10

probable cytochrome P450 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: B84733
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: B84733
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-489 <STO>
A;Cross-references: GB:AE002093; NID:g3831452; PIDN:AAC69934.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g32440
A;Map position: 2
C;Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match	15.48;	Score 394.5;	DB 2;	Length 489;
Best Local Similarity	25.68;	Pred. No. 7.3e-19;		
Matches 130; Conservative	96;	Mismatches 227;	Indels 55;	Gaps 17;

QY	9	TFLCTIVLPVLLFETAAVKLMEMLMIRVD-----PNCRSPLPFGTGLPFIGETLLQ	60
		: : : : : : : : : :	
Db	4	TGLILMWFPLIITGLFVLKW---VLKRVNWIYYSKLGEKKHYLLPBGDLGWPIVIGNMWF	60
QY	61	I-----LQRRKFLRMKRQYG--CIYKTHLEGNPTVRVMGADNVROILLGEHKLYSVQWP	113
		: : : : : : : : : : :	
Db	61	LRAKTSDPESFISQYITRYGRGTGYKAAHMFYPCVLYTTPETCRVLTDDAF-HIGWP	119
QY	114	ASVRTILGSDTLNVHGVQHKNNKKAIMRAFS-RDALEHYIPVIOQEVKSAIQEWLQKDS	172
		: : : : : : : : : : : : : : : :	
Db	120	KSTMKLIGRKSFVGISFEHKKRLRLTSAPVNGPEALSVYIQFIEETVNTDLEK*-SKMG	178
QY	173	CVLVYPEMKKLMFRIAMRILLGFEPEQIKTDEQLVEAFEEEMIKNL---FSLPIDVYFS	228
		: : : : : : : : : : : : : : : :	
Db	179	EIEFLSHLRKLTFEKVIMYIFLSSSEHV-----MDSLEREYTNLNYGVARAMGINLPGF	231
QY	229	GLYRGLRARNFIHSKIENIRKKIQDDDNENEQKYKDALQLLI---ENSRSDPEFSIQ	284

Db 232 AYHRALKARKKLVAAEQSIVTNRNRÖKÖNISSNRKMDLNDLIDVKDENGRLDD----- 286
QY 285 AMKEAATELLE-----GCHETTASTATSLVMFLGLNTEVVÖKVREEVÖEKVEMGYTPGKG 340
Db 287 ---EEIIDLLMLYNAGHESSGHLTMATILMOEHMILÖKAKEE-QÖRI-VKRAPGÖK 341
QY 341 LSMELLDÖLKYYTCVYIKETLRINPPVPGGFRVALKTEFLNGYÖIPKGMNVIYSICDTHDY 400
Db 342 LTLKETREMYLTSÖYIDETLRVITFSLTAFREAKSDVÖMDGYIIPKGWKLTFWFRNVHLD 401
QY 401 ADVEPNKEEFÖPERFMSKGLDGSRFNYPFGGSRMCVKGKEFAKVLKIFLYELLÖHCN 460
Db 402 PEIYPDPKKFDPSPRWEGYTPKAGT---FLPFGLSHLCPGNDLAKLEISIFLHFLLLKYR 458
QY 461 WILSN-GPPTMKTGPTIYPVDNLPTKFT 487
Db 459 VERSNPGCPVMFL-PHNRPKDNCLEARIT 485

RESULT 11

H96759
probable steroid 22-alpha-hydroxylase t9l24.44 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H96759
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H96759
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <STO>
A:Cross-references: GB:AE005173; NID:g11120803; PIDN:AAG30983.1; GSPDB:GN00141
C:Genetics:
A:Gene: T9L24.44
A:Map position: 1

Query Match 14.9%; Score 382.5; DB 2; Length 512;
Best Local Similarity 26.6%; Pred. No. 4.9e-18;
Matches 130; Conservative 76; Mismatches 200; Indels 83; Gaps 14;

QY 9 TFLCTIVLPLVLLFLAAVKLMWMLMIRVDPNCRSP--LPPGTMGILPFIGET----- 57
Db 19 TFLAFLITFLLAGIARRK-----RAPHRLPPGSRGWPDLIGDTFAWLNVAAG 65
QY 58 -----LÖLILÖRKKFLRMKRÖKYGCITYKTHLFGNPTVRVMGADN 96
Db 66 SHPSFEYEQIKKFVSLLCVLLILKRPDNSGFNEIRYGRIFSGSLFGKWA VVSADPDF 125
QY 97 VROILLGEHKLVSÖWPASVRTILGSDTSLNVHGÖHKNNKKAIMRAFSDALE-HY--- 152
Db 126 NRTWMÖNEKGLFÖSSYPKSFRLVGKDVITVHGÖÖRRLHSTASSMRHDÖLKTHTFLEV 185
QY 153 IPVIÖQEVKSAIÖEWLÖKDCVLYVPEMKKLMFRIAMRILGFEPEÖIKTDEÖELVEAFE 212
Db 186 IPVVMLÖTFLSNFKD---GEVVLÖDICKRVAIHLMVNÖLGVSS---SEVDEMSÖLFS 238
QY 213 EMIKNLFSLPIDVPFSGLYRGILARNFIHSKIEENIRKKIÖDDDNENEÖKYKDALÖLITE 272
Db 239 DEVDGCLSVPIDLPFTYNNKAMKARKEIRKINKTIEKRLÖ-NKAASDTAGNGVLGRILLE 297
QY 273 NSRSDPEFSLÖAMKEATELLEFGHETTASTATSLVMFLGLNTEVVÖKVREEVÖEKVEM 332

Db 298 -----EESLPNESMADFTIINLLFAGNETTSKTMFLFVYFLTHCPKAMTÖLLEE-HDRLAG 351
QY 333 GMTTPGKGLSMELLDÖLKYYTCVYIKETLRINPPVPGG-----FRVALKTEFLNGYÖIPKG 387
Db 352 GM-----LTWÖDYKTMDFTÖCVIDETFLRL----GGIAIWMREAKEDVSYÖDIYIPKG 400
QY 388 WNVYISICDTHDVADVPFNKEEFÖPERFMSKGLDGSRFN---YIPFGGSRMCVKGKEF 443
Db 401 CFVVPFLSAVHLDESYYKESLSFNPRWLDPETÖQRNMWRTSPFYCPFGGTRCPGAEFL 460
QY 444 AKVLLKIFL 452
Db 461 ARLÖIALFL 469

RESULT 12

T02263
cytochrome P450 DWARF3 - maize
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Zea mays (maize)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: T02263
R;Winkler, R.G.; Helentjaris, T.
Plant Cell 7, 1307-1317, 1995
A:Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibbe
A:Reference number: Z14648; MUID:96004534
A:Accession: T02263
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-519 <WIN>
A:Cross-references: EMBL:U32579; NID:g987266; PIDN:AAC49067.1; PID:g987267
A:Experimental source: strain B73
C:Genetics:
A:Gene: dwarf3
C:Function:
A:Description: involved in an early step in gibberellin biosynthesis
A:Pathway: gibberellin biosynthesis
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: oxidoreductase
F;325-488/Domain: cytochrome P450 homology <P45>

Query Match 14.6%; Score 375; DB 1; Length 519;
Best Local Similarity 25.6%; Pred. No. 1.6e-17;
Matches 119; Conservative 94; Mismatches 216; Indels 36; Gaps 13;

QY 41 RSPLPGTMGILPFIGETLÖLILÖRRK-----FLRMKRÖKYG--CIYKTHLFGNPTVRVMG 93
Db 69 RARLPFGEMGWPVLGGMWAFLRAFKSGKPDAFIASFVRFRGRTGVYRSEMFSSPTVLVTT 128
QY 94 ADNVRÖLLGEHKLVSÖWPASVRTILGSDTSLNVHGÖHKNNK---AIMRAFSDALE 150
Db 129 AEGCKÖVLMDDAFVT-GWPKATVALVGRSFVAMPYDEHRRIRKLTAAPINGF--DALT 185
QY 151 HYIPVIÖQEVKSAIÖEWLÖKDCVLYVPEMKKLMFRIAMRILGFEPEÖIKTDEÖELVEA 210
Db 186 GYLPFIDRIYITSLRAMADHGVSVEFTELRRMTFKIIVÖIFLG-GADÖATT--RALERS 242
QY 211 FEEMIKNLFSLPIDVPFSGLYRG-LRARNFIHSKIEENIRKKIÖDDDNENEÖKYK----- 264
Db 243 YTELNYGMRAMAINLP-GFAYRGALRRARRLLVAVLÖGVL-----DERRARAKGVSGG 294
QY 265 --DALÖLLIENSRRSDEPFSLÖAMKEATELLEFGHETTASTATSLVMFLGLNTEVVÖKV 322
Db 295 GYVMDMLRIEAÖDERGRHLDDDEIDVLYMYLINAGHESSGHITMATVFLÖENPDMFARA 354
QY 323 REEVÖEKVEMGYTPGKGLSMELLDÖLKYYTCVYIKETLRINPPVPGGFRVALKTEFLNGY 382
Db 355 KAE-QEAIMRSIPSSÖRGLTLRDRFKMEYLSÖVIDETLRVNISEVSFRÖATRDVFNNGY 413
QY 383 ÖIPKGMNVIYSICDTHDVADVPFNKEEFÖPERFMSKGLDGSRFNYPFGGSRMCVKGKE 442
Db 414 LIPKGMKVÖLWYRSVHMDPÖYVPDPTKFDPSRWEGHSPRAGT---FLAFGLGARLCPGND 470

```

QY      443 FAKVLLKIEVELLTOHCNWLNSGPPMTKTGPTIYPVDNLPTKT 487
      ||: : ||      :| : : | ||| | |
DB      471 LAKLEISVELHNFLLGYKLARTNPRCAVRVLLPHRPRVDNCLAKIT 515

```

RESULT 13

D70649
cytochrome P450 Rv3059 - Mycobacterium tuberculosis (strain H37RV)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: D70649
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: D70649
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-492 <COL>
A:Cross-references: GB:Z83866; GB:AL123456; NID:g3261691; PIDN:CAB06263.1; PID:g1781154
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3059
C:Superfamily: Mycobacterium cytochrome P450 Rv3059; cytochrome P450 homology
C:Keywords: Chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:300-461/Domain: cytochrome P450 homology <P45>
F:439/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	14.28;	Score 363;	DB 1;	Length 492;
Best Local Similarity	26.08;	Pred. No. 9.3e-17;		
Matches 129; Conservative	82;	Mismatches 210;	Indels 76;	Gaps 18;

[illegible]

RESULT 14
T06288

RESULT 14

probable cytochrome P450 T9E8.50 - Arabidopsis thaliana
N/Alternate names: protein T9E8.50
N/Contains: oxidoreductase (EC 1.-.-.-)
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 13-Aug-1999
C/Accession: T06288
R/Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Sch
submitted to the Protein Sequence Database, April 1999
A/Reference number: Z15588
A/Accession: T06288
A/Molecule type: DNA
A/Residues: 1-495 <BEV>
A/Cross-references: EMBL:AL049608
A/Experimental source: cultivar Columbia; BAC clone T9E8
C/Genetics:
A/Map position: 4
A/Introns: 126/3; 293/3; 387/3
A/Note: T9E8.50
C/Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C/Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreducta
E;295-460/Domain: cytochrome P450 homology <CYP>

Query Match	13.9%;	Score 356.5;	DB 2;	Length 495;
Best Local Similarity	26.2%;	Pred. No. 2.5e-16;		
Matches 136;	Conservative 80;	Mismatches 242;	Indels 61;	Gaps 18;

```

QY      1 MGLVTLMTVFLCTIVLPVLLFLAAVKLMWMLMIRVDPNCRSPDLPPTGMLPFIGETIQQL 60
      | | | | : | : | : | | | | | | | |
Db      1 MLITLCITLLALLKLSIKRTATKNF-----LPPSPWRLLPVIGNLHQL 46

```

```
OY      61 ILQRRKFLRMKRÖKYGCIVKTHLEGNPTVRVMGADNVRQILLGEHKLVSVOWPAS--VRT 118
        | : || : ||| : ||| : :: : ||| : | : |
Db     47 SLHTHRSLSRSLRYGPLMLLH-FGRTPVLIVSSADVADHYMKETHDLVCANRPKITKVVDK 105
```

```

QY      119 IL--GSDTLNVHGVOHKNNKR-AIMRAFSRDAL EHYIPYIOQEVKSAIOEWLÖKDSC-- 173
      11  | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      106 ILSGRDVAFAFYGEYWRÖKMSKSTICIONLNNKMVRSYEKIRKEEIKRMIEK-LEKASCSS 164

```

```
QY      174 ----VLVYPEMKKIMFRIAMRILG--FEPEQIKTDEOELVEAFEEEMIKNLFSLPIDVPF 227
        | :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db     165 SPSPVNLSQLMILTNDIICRAVALGRKYSKKDGDIVENVITFAAL---LGEFPVG EYI 221
```

QY 228 SGLYRGLRARNFIHSKIE-----ENIRKKIQQDDNENQKYKDALQLLIENSRRS 277

Db 222 PSLSWIDIRGLDH-KMEVVDKRDEFLERVVKEHEEADKFTIRSDLVDKL-LTIQSDKTG 279

```
QY 278 DEFSLOAMKEATELLFGCHETPASTATSLVMFLGINTNEVQKVREEVOEKVEMGYTP 337
      | : | : | | | : | | : | : | : | : :
Db 280 OFELKSALKLIWDMFLAGTATTSFLEWAMTEIMRNPVKVMKIOEEIRSSSPDOLLFTV 339
```

```
QY      338 GKGLSMELDQLKTYGCVIKETLRINPVPVGGF-RVAKLTFELNGYQIPKGMNVIYSI-C 395
      |      :: : | | | | | | | | | | | | : : | | | | | | | | :
Db      340 EKE-----AEKKNYLQAVIKALRPBPAPLLVPRVLSEDEVKIKGYNIPAGTQIVINAWA 394
```

QY 396 DTHDVADYFPNKEEFQPERFMSKGLE-DGSRFNYIPFGGSRMVCYKEFAKVLKIFLVE 454
| : ||| : | : |||| | : | : : |
Db 395 TORDTTWTCTDAEEFKPERHIDTNDFOGODEFKFIPEGSGKRICPGIGFTSALIGVTLAN 454

```
QY      455 LTOHCNWLISNGPP-----TMKTGPTI---YPVDNLPT 484
          : : | : |           | : : : :| :
Db      455 TVKRENMMDVFPDPOVCHDI.TEATGTIVYERKEPI.IAIPS 493
```

RESULT 15

probable cytochrome P450 T9E8.30 - *Arabidopsis thaliana*
N;Alternate names: protein T9E8.30
N;Contains: oxidoreductase (EC 1.-.-.-)

C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 13-Aug-1999
C;Accession: T06286
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schuel
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15588
A;Accession: T06286
A;Molecule type: DNA
A;Residues: 1-490 <BEV>
A;Cross-references: EMBL:AL049608
A;Experimental source: cultivar Columbia; BAC clone T9E8
C;Genetics:
A;Map position: 4
A;Introns: 129/3; 288/3; 382/3
A;Note: T9E8.30
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase
F;290-455/Domain: cytochrome P450 homology <CYP>

Query Match	13.6%;	Score 349;	DB 2;	Length 490;
-------------	--------	------------	-------	-------------

Best Local Similarity 25.0%; Pred. No. 8e-16;
Matches 127; Conservative 91; Mismatches 239; Indels 50; Gaps 19;

QY	6	LMATFLCTIVLPVLLFLAAVKLWEMLMIRVDPNCRSPPLPBGTMGLPFIGETLOILQR	65
	:	: : :	
Dd	4	LlVT-LCLTPLLALLLKSI-----LKRTTN-NNLPPSPWRLLPVIGNHQLSINTH	54
QY	66	KELRMKRÖKYGCITYKTHLFENPTVRVMGADNVRÖILLGEHKLVSVÖWPAS--VRTIL--G	121
	:	: : : : : : : : :	
Dd	55	RSLSRLSLRYGPLMLLH-FGRTPVLIVSSADVAHDILKTYDVICANRPKFVYDKILRG	113
QY	122	SDTLNSNVHGVOHKNNKK-AIMRAFSRDALEHYPVIQÖEVKSÄIQEWLÖKDSCVLYPEM	180
	:	: : : : : : : :	
Dd	114	RDVAFAPYGEYWKÖMSICIÖNLLSNKNVRSYKKIREDEIKLMIKEVENASSCSPPSPVN	173
QY	181	KLIMER----IAMRILG--FEPBQIKTDEÖELVEAFEEMIKNFLSPIDVPFSGLYRG	233
	:	: : : : : : : :	
Dd	174	LSÖLEMTLTNDIICRALGRKYSKREDGIDVENTIVRAFSAVLGE-FPIGEIIPSLSWIDK	232
QY	234	LRARNE----IHskIEENIRKKIODDDNENEÖKYKDALOILLIENSRRSDEPSLÖAMKEA	289
	:	: : : : : : : :	
Dd	233	IRGÖDHkMEEVDKRFEDEFLEERYKHEHDANKDTRSDLVDTLL--TIÖSDK---SALKLI	286
QY	290	ATELLFGGHETTASTATSIvMFGLINTEVVÖKVAREVÖEKVEMGMVTPGKGLSMELLÖL	349
	:	: : : : : : : :	
Dd	287	IWMFLAGTATSLSFLEWAMELERNPKVMKKLÖEEIRSSSRÖGLEYTEKE----AEKM	341
QY	350	KYTCGVikETLRINPPVPGGF-RVALKTfELNGYÖIPKGWNVIYST-CDTHDVADVFPNK	407
	:	: : : : : : : :	
Dd	342	DYLÖAViKEALRLRPAPLPMVPRVFSEDYTLKGYNIPAGTÖVITINAMAIQÖRDTTWGIDA	401
QY	408	EEFÖPERFMskGLE-DGSRNviPFGGSRMCVGKEFAKULLKIPLVELLÖHCNWILLSNG	466
	:	: : : : : : : :	
Dd	402	EERFERPHLDSILDÖGÖDFKFIPEGSGKRICPGIGFTSALIGVLANIVKRFNRMDVE	461
QY	467	PP-----TMKTGPTI--YpVDNLPT	484
	:	: : : : : : : :	
Dd	462	PÖRVÖHDLTEATGLVVERKPEPLIAIPS	488

Search completed: November 6, 2001, 13:35:29
Job time: 597 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:34:12 ; Search time 115.92 Seconds

(without alignments)
561.543 Million cell updates/sec

Title: US-09-668-482-2

Perfect score: 2563
Sequence: 1 MGLYTLMTFLCTIVLPVL.....GPTIYPVDNLPKRTSYVRN 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 segs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1816.5	70.9	492	13	Q9PUB4	Q9pub4 galus gall
2	1738	67.8	497	11	Q9RLF4	Q9rlf4 mus musculu
3	1646.5	64.2	492	13	Q93323	Q93323 xenopus lae
4	1034.5	40.4	512	4	Q9NR63	Q9nr63 homo sapien
5	1022	39.9	525	4	Q9NP41	Q9np41 homo sapien
6	689.5	26.9	444	2	Q59990	Q59990 synechocyst
7	598.5	23.4	477	10	Q9LVY7	Q9lv77 arabidopsis
8	532.5	20.8	485	10	Q9SJH2	Q9sjh2 arabidopsis
9	523	20.4	463	10	Q9FH76	Q9fh76 arabidopsis
10	515.5	20.1	474	10	Q9LKH7	Q9lkh7 phaseolus a
11	503	19.6	482	10	Q81077	Q81077 arabidopsis
12	503	19.6	513	10	Q64989	Q64989 arabidopsis
13	503	19.6	513	10	Q64989	Q64989 arabidopsis
14	489.5	19.1	457	10	Q65624	Q65624 arabidopsis
15	481.5	18.8	465	10	Q9FMA5	Q9fma5 arabidopsis
16	474.5	18.5	465	10	Q9LH81	Q9lh81 arabidopsis
17	460.5	18.0	478	10	Q9LH81	Q9lh81 arabidopsis
18	460	17.9	107	13	Q9PUG2	Q9pug2 galus gall
19	455	17.8	457	10	Q9M066	Q9m066 arabidopsis

20	455	17.8	524	10	Q23242	Q23242 arabidopsis
21	441	17.2	443	10	Q9LJK2	Q9ljk2 arabidopsis
22	435.5	17.0	496	10	Q9FOY4	Q9fgy4 cucurbita m
23	408.5	15.9	464	10	Q9LIC5	Q9lic5 arabidopsis
24	403.5	15.7	518	10	Q9FI38	Q9fi38 arabidopsis
25	397.5	15.5	735	10	Q9LG17	Q9lg17 arabidopsis
26	394.5	15.4	489	10	Q9ZV72	Q9zv72 arabidopsis
27	382.5	14.9	512	10	Q9FX29	Q9fx29 arabidopsis
28	370.5	14.5	504	10	Q9SNG3	Q9sng3 oryza sativ
29	366	14.3	475	2	Q9X7G9	Q9x7g9 myxococcus
30	356.5	13.9	486	10	Q9LI69	Q9li69 arabidopsis
31	356.5	13.9	495	10	Q9TOK2	Q9tok2 arabidopsis
32	353.5	13.8	474	10	Q9SHY7	Q9shy7 arabidopsis
33	350.5	13.7	491	5	Q9NGK3	Q9ngk3 tribolium c
34	349	13.6	318	10	Q9LVY3	Q9lv73 arabidopsis
35	349	13.6	490	10	Q9TOK0	Q9tok0 arabidopsis
36	347.5	13.6	516	10	Q9STI1	Q9sti1 arabidopsis
37	347	13.5	503	4	Q9HB55	Q9hb55 homo sapien
38	346	13.5	508	13	Q9PU44	Q9pu44 galus gall
39	345.5	13.5	464	10	Q04949	Q04949 arabidopsis
40	344.5	13.4	461	2	Q9K498	Q9k498 streptomyce
41	343	13.4	486	11	Q9JY3	Q9jy3 mus musculu
42	343	13.4	503	11	Q9JIP8	Q9jip8 mus musculu
43	343	13.4	511	10	Q9SWR1	Q9swr1 liguidambar
44	338.5	13.2	487	4	Q16756	Q16756 homo sapien
45	336.5	13.1	504	4	Q9HB54	Q9hb54 homo sapien

ALIGNMENTS

RESULT 1
ID Q9PUB4 PRELIMINARY: PRT; 492 AA.
AC Q9PUB4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Swindell E.C., Thaller C., Sockanathan S., Petkovich M., Jessell T.M.,
RA Eichele G.;
RT "Complementary domains of Retinoic Acid Production and Degradation in
RT the Early Chick Embryo."
RL dev. Biol. 0:0-0(2000).
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF199462; AAF09250.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450.1.
DR PRINTS; PRO0385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 492 AA; 55264 MW; 7F28B72E75C232FB CRC64;

Query Match 70.9%; Score 1816.5; DB 13; Length 492;
Best Local Similarity 70.4%; Pred. No. 1.5e-112;
Matches 345; Conservative. 63; Mismatches 79; Indels 3; Gaps 2;
QY 1 MGLYTLMTFLCTIVLPVLFLAALKVLMWMLMIRVDPCNCRSPLPGTNGLPFIGETIQL 60


```
DB 1 MGSALVASALCTFLPLFLAALVRLMDLYCASGRDPSCLPLPFGTGLPFFGETLQM 60
QY 61 ILQRRKFLMRKQKGYCYKTHLFGNPTVRVAGADNVRLILGHEKLVSVQWPASVRTL 120
DB 61 VLQRRKFLQMKRRKGYCYKTHLFGNPTVRVAGADNVRLILGHEKLVSVQWPASVRTL 120
QY 121 GSDTLNVHGVQHKNNKKAIMRAFSRDLAHEHYIPVIOEVSASIAOEWLQKDSCLVYPER 180
DB 121 GSGCLSNLHNGQHKRRKVVIMQAFSRDLQHYVPVIOEVSASIAOEWLQKDSCLVYPER 180
QY 181 KLMFRIAMRILGFEPEQIKTD-EOELVEAFEEEMIKNLSPLIDVPFSGLYRGLRARNF 239
DB 181 KLMFRIAMRILGFEPEQIKTD-EOELVEAFEEEMIKNLSPLIDVPFSGLYRGLRARNI 240
QY 240 IHSKIEENIRKTIODDNEQKYKDALQLLIENSRRSDEPFSIQAMKEATELFGGHE 299
DB 241 IHAKEENIRAKMA--RKEPEGYKDALQLLIENHTQNGEQLNMQELKESATELFGGHE 298
QY 300 TTASTATSLVMTGLNTEVQKREVEQKEMGYTPGKLSMELLDQLKYTGCVIKET 359
DB 299 TTASATSLIAFLGHLHDVLOKVRKELQGLLSGPNQEKQLMFELQKLYTGCVIKET 358
QY 360 LRINPVPVPGFRRALKTEFLNGYQIPKGNVYISICDTHDVADVPENKEEFOPERFMSKG 419
DB 359 LRSPVPVPGFRRALKTEFLNGYQIPKGNVYISICDTHDVADLPEDKDEFNDRFMSPS 418
QY 420 LEDGSRFNYIPFGGSRMCVGEKFAKVLKIFLVELTOHCNWLISNGPPTMKTGPTIYVP 479
DB 419 PEDSSRFNIPFGGSRMCVGEKFAKVLKIFLVELARSCDWQLNGPPTMKTGPIYVP 478
QY 480 DNLPTKFTSY 489
DB 479 DNLPAKFTGF 488
```

```
RESULT 2
O9RIF4 PRELIMINARY; PRT; 497 AA.
ID O9RIF4;
AC O9RIF4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME P450 RETINOIC ACID METABOLIZING ENZYME P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TOOTH;
RX PubMed=11063033;
RA Paine C.T., Paine M.L., Snead M.L.;
RT "Identification of tuftelin- and amelogenin-interacting proteins using
RT the yeast two-hybrid system.";
RL Connect. Tissue Res. 38:257-267(1998).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
EMBL: AF115769; AAD17217.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 497 AA; 56191 MW; 67CB167A997842C1 CRC64;
```

Query Match 67.8%; Score 1738; DB 11; Length 497;
Best Local Similarity 67.9%; Pred. No. 2.4e-107;

```
Matches 337; Conservative 60; Mismatches 89; Indels 10; Gaps 4;
QY 1 MGLYTLMTVFLCTIVLPLVLLAALKWEMLMIRVDPNCRSPPLPPTGMLPFIGETLQL 60
DB 1 MGLPALLATALCTFVLLPLFLAALKWLDLYCVSSRDRSCALPLPPTGMPFFGETLQM 60
QY 61 ILQRRKFLMRKQKGYCYKTHLFGNPTVRVAGADNVRLILGHEKLVSVQWPASVRTL 120
DB 61 VLQRRKFLQMKRRKGYCYKTHLFGNPTVRVAGADNVRLILGHEKLVSVQWPASVRTL 120
QY 121 GSDTLNVHGVQHKNNKKAIMRAFSRDLAHEHYIPVIOEVSASIAOEWLQKDSCLVYPER 176
DB 121 GSGCLSNLHDSHKKQKRYIMQAFSRDLQCYLVIAEVSCLQEWL---SCGERGLV 177
QY 177 YPEMKKLMFRIAMRILGFEPEQI--KTDEQELVEAFEEEMIKNLSPLIDVPFSGLYRGL 234
DB 178 YPEVKRLMFIAMRILGCEPGPAGGEDEQQLVEAFEEEMIRNLSPLIDVPFSGLYRGV 237
QY 235 RARNFIHSKIEENIRKTIOD--DDNENEQKYKDALQLLIENSRRSDEPFSIQAMKEATEL 293
DB 238 KARNLIHARIEENIRAKIRLQATEPDDGCKDALQLLIHSWGERGLDMQALKOSSTEL 297
QY 294 LFGHETTASTATSLVMTGLNTEVQKREVEQKEMGYTPGKLSMELLDQLKYTG 353
DB 298 LFGHETTASATSLTYGLYPLVLOKVRKEIKSKGLLCKSNQDNKLMETLQKYIG 357
QY 354 CVIKETLRINPVPVPGFRRALKTEFLNGYQIPKGNVYISICDTHDVADVPENKEEFOPE 413
DB 358 CVIKETLRINPVPVPGFRRALKTEFLNGYQIPKGNVYISICDTHDVADLPENKEEFPD 417
QY 414 REMSKLEDGSRFNYIPFGGSRMCVGEKFAKVLKIFLVELTOHCNWLISNGPPTMKTG 473
DB 418 RFTVPHPEDASRFNIPFGGSRMCVGEKFAKVLKIFLVELARHCDWQLNGPPTMKTG 477
QY 474 PTIYVVDNLPTKFTSY 489
DB 478 PTIYVVDNLPTKFTSY 493
```

```
RESULT 3
O93323 PRELIMINARY; PRT; 492 AA.
ID O93323;
AC O93323;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen J.L., Grunz H., Panitz F., Pieler T., Hollemann T.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
EMBL: AF057566; AAC25158.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 492 AA; 55459 MW; D1D4B87651BF2D3E CRC64;
```

Query Match 64.2%; Score 1646.5; DB 13; Length 492;
Best Local Similarity 63.2%; Pred. No. 2.8e-101;
Matches 312; Conservative 84; Mismatches 87; Indels 11; Gaps 4;

[illegible]

	Best Local Similarity	42.38;	Pred. No. 1e-60;	Mismatches 168;	Indels 17;	Gaps
QY	1	MGLYTLMTVFLCTIVLPVLLFLA	AVKLWMMLMIRVDNCRSP	LRPGTGMGLPIGETLQL	60	
Db	6	LDLVSA	LATLACLAVSVTL	LAVSQQLWRMA	TRKSCKLP	IPKGS
QY	61	ILQRKRLMKRKQYGC	IKYKTHLFGNP	TVRVMGADNV	RIILG	EHKLVSYQWPA
Db	66	LLQSGFQSSRREKYGN	VFKTHLGRPLIRVTGA	ENVRKILMG	EHHLVST	EMPRSTRMLL
QY	121	GSDTL	SNVHGVOHKNNKKAI	MRAFSRDALEHYIP	VIQOEYKSAIQEW	LQKDSCLVYPEM
Db	126	GPNTV	SNISIGDIHRNKRK	VFSKIF	SHALESYLPKIQ	LVYDTLRAMSSHP
QY	181	KKLMFR	IA	MRIL	LGFE-PEQIK	TDEQLVEAFEEMIK
Db	186	QKLF	FRMAIR	LLGFSI	PEE--DIGH	LEFVYQO
QY	240	IHSKIE	ENIRKKIQDDN	ENENQYKDALQ	LIENSRRSD	EPESLOAMKEA
Db	243	LQKLE	KAIREKIQCTQ	GKD--YLD	ADLLI	ESSKEHGKEMQEL
QY	300	TTAST	ATSLV	WFLG	INT	EVQKVREVOEK
Db	300	TTAS	ASTSLIMQ	LKHP	TVLEKLR	DELRAH--GILHSGGC
QY	355	VIKET	LIRINP	PVGE	FRVALK	T
Db	357	VIK	EV	MR	LFTPI	SGGYRTVLQ
QY	415	FMSK	GLED-GSR	FNI	IPFGG	SRMCVGR
Db	417	FSQ	ARSEDK	DGRFHY	LPFGG	GVRTCLGK
QY	473	GPTI	YPVD	NLP	PTKF	486
Db	477	VP	VLHP	VDGL	SVKF	490

```

RESULT      5
O9NP41
ID      O9NP41      PRELIMINARY;      PRT;      525 AA.
AC      O9NP41;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT      01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE      WUGSC:H_NH0493JL16.1 PROTEIN.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99063792; Pubmed=9847074;
RA      Sulston J.E., Waterston R.;
RT      "Toward a complete human genome sequence.";
RL      Genome Res. 8:1097-1108(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Bourne S., Bauer C., Pape K., Jones T.;
RT      "The sequence of Homo sapiens BAC clone RP11-493J16.";
RL      Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RA      Waterston R.;
RL      Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC      -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC      OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC      SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC      EMBL; AC007002; AAF65576.1; -.

```

DR InterPro; IPR001128; -
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 525 AA; 59124 MW; 71D47B6752A60315 CRC64;

Query Match 39.9%; Score 1022; DB 4; Length 525;
Best Local Similarity 41.2%; Pred. No. 7.1e-60;
Matches 209; Conservative 101; Mismatches 167; Indels 30; Gaps 9;

QY 1 MGLYLMVTFLLCTIVLPVLLFLAAVKLWEMLMIRVDNCRSPLPPTGMLPFIETQL 60
DB 6 LDVLSALATLAACLVSVTLTLLAVSQQLWLRNATRDCKSLPIPKSGMGFPLIGETGM 65
QY 61 ILQR-----RKFLRMKROKYGCIYKTHLFGNPTVRVMGADNVRLGEGHKL 107
DB 66 LLQKCTLRTRVWLPQSGSGFQSSRRKYGWFKTHLLGRPLIRVTGAENVRKILMGEHHL 125
QY 108 VSVQWPASVRTILGSDTLNSNVHGVQHKKKKAIMRAFSRDALSHYIPVIOQEVKSAIQEW 167
DB 126 VSTWPRSTRMLGSPNTVNSIGDIHRNKRKFSKIFSHALLESYLPKIQLVITQDLRAW 185
QY 168 LQKDCVLYPEMKIMRIAMRILGFE-PEQIKTDEQELVEAFEMIKNLFSLPIDVP 226
DB 186 SSHPEALINVQEAQKLFERMAIRVLLGFSIPEE--DLGHLFEVYQQFVDNVFSLPVDLP 242
QY 227 FSGLYGLRARNFIHSKIEENIRKKIQQDDNENQKKDALQLLIENSRRSDEPSLQAM 286
DB 243 FSGYRRGIQARQILOKLEKAIKREKIQCTOGKD--YLDALDLIESSKEHGKEMTQEL 299
QY 287 KEATELLFGGHETTASTATSLVMFLGINTENVQKREEVQEKVEMGMY---TPGKG-L 341
DB 300 KDGTELELFAAYATFASASTSLIMQLKHPVLEKLRDELRAH--GILHSGGCPCEGTL 356
QY 342 SMELLDQKLYTGCIVIKETLRINRPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVA 401
DB 357 RLDTLGRLYLDCAVKEWMLFTPISGYRVVLQTFELDGFQIPKGSVMYSIRDTHDTA 416
QY 402 DVFPNKEEFQPERFMSKGLD-GSRFNYIPFGGSRMCVGKEFAKVLKIFLVELTOHCN 460
DB 417 PVFKDVNVFDPDRFSQARSEDKDGRFHYLPFGGVRCTLGKHLAKLFLKYLAVELASTSR 476
QY 461 WILSNGP-PTMKTGPTIYPVDNLPTKF 486
DB 477 FELATRTFPRTITLVPVLPVLDGLSVKF 503

RESULT 6
Q59990 PRELIMINARY; PRT; 444 AA.
AC Q59990;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE PUTATIVE CYTOCHROME P450 120.
GN CYP120 OR CYP OR SLR0574.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA Tabata S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiyama M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).

RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; D64003; BAA10496.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; p450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Hypothetical protein; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 444 AA; 50578 MW; 8F62A9ED3B54BDC CRC64;

Query Match 26.9%; Score 689.5; DB 2; Length 444;
Best Local Similarity 34.5%; Pred. No. 6e-38;
Matches 156; Conservative 99; Mismatches 170; Indels 27; Gaps 10;

QY 43 PLRPGTGLPFIETQLQIRKFLRMKROKYGCIYKTHLFGNPTVRVMGADNVRL 102
DB 12 PIRPGDFGLPWLGETLN-FLNDGDFGKKRQGFPIFKTRLFGKNVIFISGALANRFLFT 70
QY 103 GEHLVSVQWPASVRTILGSDTLNSNVHGVQHKKKKAIMRAFSRDALSHYIPVIOQEVKS 162
DB 71 KEQETPQATWPLSTRILLGNALATQMGIEHRSRKLQYQAFLPRTLDSTLPKMDGIYOG 130
QY 163 AIOEWLQKDCVLYPEMKIMRIAMRILGFEPEQIKTDEQELVEAFEMIKNLFSLP 222
DB 131 YLEQW-GKANEVIWYPOLRRMFDVAATLFMG--EKVSQNPQ-LFPWFETIYIQLFSLP 185
QY 223 IDVPFSGLYGLRARNFIHSKIEENIRKKIQQDDNENQKKDALQLLIENSRRSDEPS 282
DB 186 IPLPNTLFGKSQARALLAELEKIKARQQPSE-----EDALGILLAARDNNQPLS 240
QY 283 LQAMKEATELLFGGHETTASTATSLVMFLGINTENVQKREEVQEKVEMGMYTPGKGLS 342
DB 241 LPELKDOIILLFAGHETLTSALSSFCLLGQHSIDIREVRQOE-QNKIQL-----SQELT 294
QY 343 MELLDQKLYTGCIVIKETLRINRPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVA 402
DB 295 AETLKMPYLDQVLEVLRLIPVGGGFRRELIDCCQGFHFPKGLVSYQISQTHADPD 354
QY 403 VFPNKEEFQPERFMSKGLDGSR----FNYPFGGSRMCVGKEFAKVLKIFLVELTO 457
DB 355 LYPDEKFDPERF----TPDGSATHNPPFAHVFPFGGLRECLGKEFAKLEMLFATRLIQ 410
QY 458 HCNWITSNGPP-TMKTGPTIYPVDNLPTKFTS 488
DB 411 QFDWTLIPGQNLLELVTPSPRPKDNLRVKLHS 442

RESULT 7
Q9LVY7
ID Q9LVY7 PRELIMINARY; PRT; 477 AA.
AC Q9LVY7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE CYTOCHROME P450-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:31-63(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AB018112; BAA96885.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 477 AA; 54850 MW; 304B4B2C4970E405 CRC64;

Query Match 23.4%; Score 598.5; DB 10; Length 477;
Best Local Similarity 31.0%; Pred. No. 7.1e-32;
Matches 155; Conservative 94; Mismatches 182; Indels 69; Gaps 18;

QY 14 IVLPVLLFLAIVKIMLMIRVDPNCSP-LPPGTMGLPIGTLQILQRR-----KF 67
DB 3 MAIMILFLSSILSLLLRLKHLSHESYFNLPPGNTGLPLIGSFSFLSAGRGHPKF 62
QY 68 L--RMKR--QKYGCIYKTHLFGNPTVRVWGADNVRQILLGHEKLVSVQWPASVRTIGSD 123
DB 63 ITDRVRRFSSSSSCVFETHLFGSPTAVVTGASGNKFLFTENKLVYSWWPDVSNKIPSS 122
QY 124 TUSNVHGVQHKKKKAIMRAFSR-DALEHIPIVIOEVKSAIO-EWLKDSCLVYPENK 181
DB 123 MQTS--SKEARKLMLLSQFMKPEALRRYGVMEIAQRHFTETWANQDQ-VIVFPLTK 179
QY 182 KIMERIAMRILGFE-PEQIKTDEQLEVAEEMIKNLFSLPIDVPSGLYRGLRANFI 240
DB 180 KFTFSIACRSFLSMEDPARVR---QLEEQFNTVAVGIFSIPIIDLPGTRFNRAIKASRL 235
QY 241 HSKIEENIRKKIQQDDNENQKYKDALQ-----LLIENSRRSDEPFSLOAMKEATEL 293
DB 236 RKEVSAIVRQ-----KEELKAGKALEHEDILSHMLNIGETKDED-----LADKIIGL 284
QY 294 LFGGHETTASTATSLVFLGLNTEVVOQVREE---VOEKVEMGYTPGKLSMELLDOL 349
DB 285 LIGGHDTASTVCTVYVNYLAEFPHYQRYLQEOKEILKEKKEK-----EGLRWEDIEKM 338
QY 350 KYTGCVIKETLRINPVPVGGFRVALKTEFLNGYQIPKGMNVIYSICDTHDVADVPEPKKE 409
DB 339 KYSWNVACEVMRIYPLSGTFREAIHFSFKGYIPKWKLYMSATATTHMNPDIYFPEPER 398
QY 410 FQPERFMSKGLDGSRENYIPFGGSRMCVGEFAKVLKIFLVELLTHQCNWILSNPPT 469
DB 399 FEPNRFEGSGPKP---YTYVFGGGPRCQPKKEYARLEILIFMHNLVNRFKW----- 447
QY 470 MKTGPTIY-----VDNLP 483
DB 448 ----EKVFPNENKIYVDLP 463

RESULT 8
Q9SJH2 PRELIMINARY; PRT; 485 AA.
AC Q9SJH2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE CYTOCHROME P450.
GN AT2G42850.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana."
RL Nature 402:761-768(1999).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AC006931; AAD21724.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 2.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 485 AA; 55405 MW; 08B0B16474620F82 CRC64;

Query Match 20.8%; Score 532.5; DB 10; Length 485;
Best Local Similarity 28.2%; Pred. No. 1.7e-27;
Matches 143; Conservative 101; Mismatches 214; Indels 49; Gaps 15;

QY 4 YTLMTVFLC--TIVLPVLLFLAIVKIMLMIRVDPNCSP-LPPGTMGLPIGTLQILI 61
DB 7 FVLSWVFLCIAATISSTLFFER--KKNHRTTKRIQK--KKLLPGEMGLPIGETMDY 62
QY 62 LQRRK-----FLRMKRQKYGCIYKTHLFGNPTVRVWGADNVRQILLGHEKLVSVQWPAS 115
DB 63 KAQKSNRVEDFVNPRIRIKHGNIKFTKTRIMGSPTIYVNGAEANRLILSNESLVVSSWSS 122
QY 116 VRTILGSDTSLNVHGVQHKKKKAIMRAFSRDALAEHIPIVIOEVKSAIO-EWLKDSCLV 174
DB 123 SVOLGMNCINAKQGEKHRVLRGIYANSLSYIGLESILPKCDYVKEFHETWGRKEE-I 181
QY 175 LVYPEMKIMERIAMRILGFEPEQIKTDEQLEVAEEMIKNLFSLPIDVPSGLYRGL 234
DB 182 SLYRSKAVLFTTVFECLYG-----IKV-EIGMLEVERVLEGVFALPVEPQSKFARAK 235
QY 235 RARNEIHSKIEENIRKKIQQDDNENQKYKDAL-QLLIENSRRSDEPFSLOAMKEATE- 292
DB 236 KARLEIETFLVGKVRERKREMEKGAEPNTTLFSRLVE-----ELIKGVITTE 284
QY 293 -----LFGGHETTASTATSLVFLGLNTEVVOQVREEVOEKVEMGYTPGKLSM 343
DB 285 EVVDNMVLLVFAAHDTSYAMSMTEKMLAQHPTCRDPLQLQHAQIKANKG---EGEYLT 341
QY 344 ELLDQKYYTCVIKETLRINPVPVGGFRVALKTEFLNGYQIPKGMNVIYSICDTHDVADV 403
DB 342 EDVKKMKYSWQVVRRETMRSLPIFGSFRKAVADIDYGGTYIPKWKILMTYGTHTYNP 401
QY 404 FPNKEEFQPERFMSKGLDGSRENYIPFGGSRMCVGEFAKVLKIFLVELLTHQCNWIL 463
DB 402 FQDPMSPDPTNF-DKPIQ--ATYLLPFGGPRLCAGQLAKISILVFMEHYVTGFDWSL 457
QY 464 SNGPPTMKTGPTIYVDNLPFTKTSYV 490

Db 458 VYFDETISMDPLFPFSLGMPKISPKV 484

RESULT 9

Q9FH76 PRELIMINARY; PRT; 463 AA.

AC Q9FH76;

DT 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)

DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)

DE CYTOCHROME P450.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RX MEDLINE=20181125; PubMed=10718197;

RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H., Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones."

RL DNA Res. 7:31-63(2000).

DR EMBL: AB020744; BAB10255.1;

SQ SEQUENCE 463 AA; 52366 MW; CCD17293F553F812 CRC64;

Query Match 20.4%; Score 523; DB 10; Length 463;

Best Local Similarity 29.4%; Pred. No. 6.8e-27;

Matches 145; Conservative 93; Mismatches 203; Indels 52; Gaps 15;

QY 2 GLVTLM--VTFLCTIVLPLVLLAAYKLMWMLIRVDPNCRSPLPPTGMLPFIGETL 58

Db 5 GLFLTLAALFLC-----LLRFIAGV-----RRSSTKLPPLPGTMGYPYVGETF 50

QY 59 QULIQ-RRKFLRMKROKYGCIYKTHLFGNPVYRVGADNVRQILLGHEKLVSVQWPASVR 117

Db 51 OLYSDPNVFFAAKORRGYSVFETHVLGCPVMISSPEAKFVLVTKSHLFKPTFPASKE 110

QY 118 TILGSDTLNVHGVQHNKKKAIMRAFSRDALSHYIPVIOQEVKSAIQEWLQKDSCLV-V 176

Db 111 RMLGQAIFFHQGDYHSLKRLVLRAFMPDAIRNMVPHIESIAQESLSN- --DGTQLNT 167

QY 177 YPEMKLMFRIAMRLLGFEPEQIKTDEQELVEAFEEIMKNLFSLPIDVPSGLYGLRA 236

Db 168 YQEMKTYTFNVALISLIGKDEVYRED--LKRQYIILEKGYNSMPINLPGLTFHKAMKA 224

QY 237 RNFHSKIEENIRKKIODDNEQKYKDALQLLIEN-SRSDEPFSIQAMKEATELLF 295

Db 225 RKEL-AQILANILSK----RRQNPSSHTDLGSEMEDKAGLTDEQIA-----DNIIGVIF 274

QY 296 GGHETTASTATSLVMEFLGLNTEVVQKREEV---QEKVEMGMVTPGKLSMELLDDQKY 351

Db 275 AARDTASVLTWLKYLADNPVLEAVTEEQMATRKDKKE-----GESLTWEDTKKMP 328

QY 352 TGCVIKETLRINPVGGRVALKTFELNGYQIPKGMNVISICDTHVADVFPPNKEEFQ 411

Db 329 TYRVIOETLRAATILSFTFRREAVEDVEYEGYLIPKGWKVLPLFRNIHNADIFSDPGKF 388

QY 412 PERFMSKGLDGSRFN-YIPFGGSRMCVGEKFAVLLKIFLVLELTQHCHWILSNGPPTM 470

Db 389 PSRF-----EVAPKPNFMFPGSGIHSCPGNELAKLEISVLHHLTTKRWISIVGPSGI 443

QY 471 KTGPTIYPVDNLP 483

Db 444 QYGFALLPQNGLP 456

RESULT 10

Q9LKH7

ID Q9LKH7 PRELIMINARY; PRT; 474 AA.

AC Q9LKH7;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)

DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)

DE CYTOCHROME P450.

GN CIPCP.

OS phaseolus aureus (Mung bean) (Vigna radiata).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;

OC Fabales; Fabaceae; Papilionoideae; Vigna.

OX NCBI_TaxID=3916;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. 2937;

RA Yang M.T., Chen Y.M.;

RT "Cloning and sequencing of a Vigna radiata cDNA encoding cytochrome P450."

RT Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.

CC -I- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH + OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).

CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY SIMILARITY).

CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC EMBL: AF279252; AAF89209.1; -.

DR InterPro; IPR001128; -.

DR Pfam; PF00067; P450; 2.

DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.

KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome; Monooxygenase; Oxidoreductase.

KW SEQUENCE 474 AA; 54037 MW; A2542A809C5BAC6D CRC64;

Query Match 20.1%; Score 515.5; DB 10; Length 474;

Best Local Similarity 29.4%; Pred. No. 2.2e-26;

Matches 141; Conservative 99; Mismatches 209; Indels 31; Gaps 15;

QY 14 IYLPVLLFLAAYKLMWMLIRVDPNCRSPLPPTGMLPFIGETLQILQRRK-----FL 68

Db 2 VSLPFLLLFLAASA-AAIFLHRFAFSRRKRLPSPSYGLPFIGETLQILASAYSSNPEPFM 60

QY 69 RMKROKYGCIYKTHLFGNPVYRVGADNVRQILLGHEKLVSVQWPASVRTILGSDTLNV 128

Db 61 DERVRRYSGIFMTHVFGEPTFSADPELNRFILQNEKRLDSCYPSISNLGKHSLLM 120

QY 129 HGVQHNKKKAIMRAFSRDAL--EHYIPVIOQEVKSAIQEWLQKDSCLVYPEMKLMFR 186

Db 121 KGALHR-RMHSLTMSFANSITKDHLHIDRLGLNDTWSDR--VTLMQAKKITFE 176

QY 187 IAMRLLGFEPEQIKTDEQELVEAFEEIMKNLFSLPIDVPSGLY-RGLRARNFIHSKIE 245

Db 177 LTVKQILMSFDPDEM--TESLRKEVYLVIEGFTLPLPL-FSTTYRAIKARKVAEALT 232

QY 246 ENIRKKIODDNEQKYKDALQLLIENSRRSDEPFSIQAMKEATELLFGGHETTASTA 305

Db 233 LVVRQR-REIYNQGEKESDMIGALL---ASGDHESDDQIVDFLLALVAGYETTSTIM 287

QY 306 TSLVMEFLGLNTEVVQKREVEQEVEMGMVTPGKLSMELLDDQKYTGCVIKETLRINP 365

Db 288 TLAVKFLTEPLALQKEE-HDQIR-ARSDPGAPLEMTDYKSMVFQHYVETLRVANI 345

QY 366 VPGGRVALKTFELNGYQIPKGMNVISICDTHVADVFPPNKEEFQPERFMSKGLDGSR 425

Db 346 IGGIFRRAATTDIDIKGYTIPKGMKVFAFRAVHLNPEYTKDARTFNPWRQSSSEANP 405

QY 426 FN-YIPFGGSRMCVGEKFAVLLKIFLVLELTQHCHWILSNGP-----PTMKTGPTIYPV 479

Db 406 ANYTTPFGGPRCLCPGYELARVYLVFLHRIYTRFSWVPAEDKLVFFPTTRQKR-YPI 464

RESULT 11

081077
ID 081077 PRELIMINARY; PRT; 482 AA.
AC 081077;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PUTATIVE CYTOCHROME P450.
GN T914.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC T914 genomic sequence."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AC005315; AAC33235.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; p450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 482 AA; 55175 MW; AB6A07AB2778DD3B CRC64;

Query Match 19.6%; Score 503; DB 10; Length 482;
Best Local Similarity 28.9%; Pred. No. 1.5e-25;
Matches 140; Conservative 92; Mismatches 224; Indels 28; Gaps 11;

QY 7 MVTFLCTIVLPVLLFLAAYKLMEMLMIRVDPNCSPPLPGTMGLPFIGETLQILQIR-R 65
DB 22 LITLTIWVVVVLLF---KWLHWKEQRL-----RLPGSGMGLPYIGETLRLTYTENPN 71
QY 66 KFLRMKROKYGCIYKTHLFGNPTVRVMGADNVROILLGEHKLVSQVMPASVKTILGSDTL 125
DB 72 SFQATRONKYGDIFKTHILGCPVMISSPEARMVLSKAHLFKPTYPSPSKERMIGPEAL 131
QY 126 SNVHGVQHKKKKAIMRAFSRDALSHYIPVIOQEVKSAIQEWL-QKDCVLVYPEMKKLM 184
DB 132 FEHQGPYHSTLKRIVQSSFMPSALRPTVSHIELLVQLTSSWTSQKSINTLEY--MKRYA 189
QY 185 FRIAMRILGFEPQIKTDEQELVEAFEEIMIKNLFSLPIDVPFSGLYRGRLARNFTHSKI 244
DB 190 FDVAIMSAFGDKKEPTTIDVIKLL--YQRLERGVNSMPLDPLGTLFHKSMKAR---IEL 243
QY 245 EENIRKKIQDDNENEQYKDALQLLIENSRRSDEPFSLQAMKEATELLFGHETTAST 304
DB 244 SEELRKVI-EKRRENGREEGGLLVLLGAKDKRNGLSDSQIADNIIGVIFAATDTTASV 302
QY 305 ATSLVMFLGLNTEVNVQKREEVQEKVEMGMVTPGKLSMELLDQLKYTGCVIKETLRINP 364
DB 303 LFWLLKYLHDHPNLLQEVSR-QFSIRQIKKENRKRISWEDTRKMPLTRVIOETLRAAS 361
QY 365 PYPGGFRVALKTFELNGYQIPKGMNVYISICDTHVADVFPNKKEEFQPERMSKGLDGS 424
DB 362 VLSFTFRAVQDVEYDGYLIPKGWVPLPLFRRIHHSSEFFPDPEKFDPSRF-----EVAP 416
QY 425 R-FNYIPFGGSRMCVGEFAKVLKIFLVLELTHQCNWILSNGPPTMTKTGPTIYPVDNLP 483
DB 417 KPYTYMPFGNGVHSCPGSELAKLEMLILLHLLTTSFRWEVIGDEEGIQGPPVPPKGLP 476
QY 484 TKFT 487
DB 477 IRYT 480

RESULT 12
ID 064989 PRELIMINARY; PRT; 513 AA.

AC 064989;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE STEROID 22-ALPHA-HYDROXYLASE.
GN DWF4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WS-2;
RA Choe S., Dilkes B.P., Fujioka S., Takatsuto S., Sakurai A.,
RA Feldmann K.A.;
RL Plant Cell 10:0-0(1998).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF044216; AAC05093.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; p450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58867 MW; B1639BDD9A4DA6F3 CRC64;

Query Match 19.6%; Score 503; DB 10; Length 513;
Best Local Similarity 27.6%; Pred. No. 1.7e-25;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

QY 4 YTLMTFLCTIVLPVLLFLAAYKLMEMLMIRVDPNCSPPLPGTMGLPFIGETLQILI-- 61
DB 7 HTLLPLLLPSLSSLFL-----ILKRRNRKTRFNLPFGKSGWPFLEGTIGYLLKP 58
QY 62 ---IQRRKFLRMKROKYGCIYKTHLFGNPTVRVMGADNVROILLGEHKLVSQVMPASVRT 118
DB 59 YTAATLGDQMCOHVSQYKGIYRSNLFGEPTIVSADAGLNRFILQNEGRLEFECYPRSIGG 118
QY 119 ILGSDPTSNVHGVQHKKKKAIMRAFSRDALSHYIPVIOQEVKSAIQEWL-QKDCVLVY 172
DB 119 ILGKWSMLVLVGDMDHRD---MRSISLNFSLSHARLRTILKDVERTLFLVDSW-QQNS 172
QY 173 CVLVYPEMKKIMFRIAMRILGFEPQIKTDEQELVEAFEEIMIKNLFSLPIDVPFSGLYR 232
DB 173 IESAQDEAKKFTENLMAKHMSDGEET--EQLKKEYVTFMKGVSAPLNLPGTAYHK 230
QY 233 GLRAR-----NFIHSKIEENIRKKIQDDNENEQYKDALQLL-----IENSRRSDE----- 279
DB 231 ALQSRATLIKFERKMEERKLDIKEDQEEVEKTEDEAEMSKSDHVRKQRTDDDLGWV 290
QY 280 ----PFSLQAMKEATELLFGHETTASTATSLVMFLGLNTEVNVQKREE---VQEKVE 331
DB 291 LKHSNLSTEOILLDLISLFFAGHETSSVAIALAIFFLQACRAVEELREHLEIARAKE 350
QY 332 MGMTTPGKGLSMELLDQLKYTGCVIKETLRINPPVPGGFRVALKTFELNGYQIPKGMNVY 391
DB 351 LG---ESELNWDYDKKMDFTQCIVINETLRLGNVVRFLHRKALKDVRYKGYDIPSGWVL 406
QY 392 YSICDTHVADVFPNKKEEFQPERFM-----SKLEDGSRF--NYIPFGGSRMCVGE 442
DB 407 PVISAVHLDSGRYDQPNLFNPNRWQOQNGASSSGSFSFTWGNMYMPFGGGRPLCAGSE 466
QY 443 FAKVLKIFVLELTHQCNWILSNGPPTMTKTGPTIYPVDNLP 483
DB 467 LAKLEMAVFIHLLVLKFNWELAEDDQ-----PFAFPVDFP 502

RESULT 13
ID 09SCQ9 PRELIMINARY; PRT; 513 AA.
AC 09SCQ9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE STEROID 22-ALPHA-HYDROXYLASE (DWF4).
GN T3A5.40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bioecker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AL132979; CAB62435.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58867 MW; B1639BDD9A5D7C93 CRC64;

Query Match 19.6%; Score 503; DB 10; Length 513;
Best Local Similarity 27.6%; Pred. No. 1.7e-25;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

QY 4 YTLWTFPLCTIVLPVLFLLAAVKLWEMIMIRVDPNCRSPLPPGTMGLPFIGETLQLI-- 61
Db 7 HTLLPLLLPSLLSLLLFL-----ILLKRRNRKTRFNLPPGKSGWPFLGETIGYLKP 58
QY 62 ---LQRRKFLRMKRQKYGCITYKTHLFGNPVTRVMGADNVRLIGEHKLVSVQWPASVRT 118
Db 59 YTAATTLGDFMQOHVSKYGYKTYRSNLFGEPTIVSADAGLNRFILQNEGRLEFECSPRSIG 118
QY 119 ILGSDTSLNVHGVQHKKKKAIMAFSRDALEH-----YIPVIOQEVKSAIQEWLQKDS 172
Db 119 ILGKWSMLVLVGDMDRD---MRSISLNLFLSHARLRTLLKDYERHTLFLVLDWS-QQNS 172
QY 173 CVLVYPEMKKLMFRIAMRILLGFEPEQIKTDEQELVEAFEEIMKNLFSLPIDVPFSGLYR 232
Db 173 IFSAQDEAKKFTFNMAKHIMSDPGEET--EQLKKEYTFMKGVVASPLNPGTAYHK 230
QY 233 GLRAR---NFIHSKIEENIRKKIODDNNENQKYKDALQL---IENSRSD----- 279
Db 231 ALQSRATILKFIERKMEERKLIDKEQDEEEVKTDEAEEMSKSDHVRKQRTDDDLGW 290
QY 280 ----PESLQAMKEAATELLFGGHETTASTATSLVMFLGLNTEVVQKVREE---VQEKVE 331
Db 291 LKHSNLSSTEQILDILSLFAGHETSSVAIALAIFFLQACPKAVEELREHLEIARAKE 350
QY 332 MGMTTPGKGLSMELLDQLKYYCIVIKETLRINPVPVPGGFRVALKTFELNGYQIPKGVNI 391
Db 351 LG----ESELNWDYKKMDFTQCVINETLRIGNVVRFLHRKALKDVRKYKYDIPSGWKYL 406
QY 392 YSICDTHDVADVFPNKEEFQPERFM-----SKGLEDSRF--NYIPFGGSRMCYGE 442
Db 407 PVISAVHLDNSRYDQPNLEFPWRWQOQNNGASSSGSGSFSTWGNMYMPFGGPRLCAGSE 466
QY 443 FAKVLKIFLVELTQHCNWLISNGPPTMKTGPTIYPVDNLP 483
Db 467 LAKLEMAVFIIHLVLKFNWELAEDDK-----PFAFPYVDPP 502

RESULT 14
ID 065624 PRELIMINARY; PRT; 457 AA.
AC 065624;
DT 01-AUG-1998 (Tremblrel. 07, Created)

DT 01-AUG-1998 (Tremblrel. 07, last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE CYTOCHROME P450.
GN T18B16.200 OR AT4G19230.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-457 FROM N.A.
RA Van Der Schueren J., Chuang Y.J., Aert R., Defoor E., Robben J.,
RA Volckaert G., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AL021687; CAA16713.1; -.
DR EMBL; AL161550; CAB78925.1; -.
DR HSSP; P33006; ICPT.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 457 AA; 52436 MW; 6105FB7C181E4F07 CRC64;

Query Match 19.1%; Score 489.5; DB 10; Length 457;
Best Local Similarity 28.2%; Pred. No. 1.1e-24;
Matches 127; Conservative 97; Mismatches 183; Indels 43; Gaps 12;

QY 20 LFLAAVKLN--EMLMIRRVDPNCRSPLPPGTMGLPFIGETLQLILQ-RRKFLRMKRQKYG 76
Db 10 LFAGSLFLYLRLCLISQRRFGSSKLPPLPGTMGWYVGETFQLYSQDPNVFFQSKQKRYG 69
QY 77 CIYKTHLFGNPVTRVMGADNVRLIGEHKLVSVQWPASVRTILGSDTSLSNVHGVQHKNK 136
Db 70 SVFKTHVLGCPVMISSPEAKFVLVTKSHLKFPTFPASKERMLGQAIFFHQGDYHAKL 129
QY 137 KKAIMAFSRDALEHYIPVIOQEVKSAIQEWLQKDSCLVYPEMKKLMFRIAMRIILGFE 196
Db 130 RKLVLRAFMPEISIRNMVDPDIESIAQDSLRSW--EGTMINTYQEMKTYTFNVALLSIFG-- 185
QY 197 PEQIKTDE---QELVEAFEEIMKNLFSLPIDVPFSGLYRGLRARNFHSKIEENIRKRI 252
Db 186 -----KDEVLYREDLKRCRYILEKGYNMPPVNLPGTLFHKSMKAR-----KELSQILARI 235
QY 253 QDDNENENQKYKDAL-QLLIENSRSDPEPSLQAMKEAATELLFGGHETTASTATSLVMF 311
Db 236 LSERQNGSSHNDLLGSEFMGKRELLTDEQIA----DNIIGVIFAARDTTASVMSWILKY 290
QY 312 LGLNTFVVQKVREEV---QEKVEMGYTPGKGLSMELLDQLKYYTCVIKETLRINPVP 367
Db 291 LAENPNVLEAVTEEQMAIRKDEE-----GESLTWGDTRKMPLTSRVIGQETLRVASILS 344
QY 368 GGFRRVALKTFELNGYQIPKGVNIYSICDTHDVADVFPNKEEFQPERFMSKGLEDSGRFN 427
Db 345 FTFRRAVEDVEYEGYLLIPKGVNLPFRNIHHSADIFSNGKFDPSRF-----EVAAPKPN 399
QY 428 -YIPFGGSRMCYGEFAKVLKIFLVELT 456

```

Db      400 TEMPEGNGTHSCPGNELAKLEMSIMIHLLT 429
          ::||| | : | | | | : : | : ||

```

Search completed: November 6, 2001, 13:38:25
Job time: 253 sec

RESULT	15
09FMA5	
ID	09FMA5 PRELIMINARY; PRT; 465 AA.
AC	09FMA5;
DT	01-MAR-2001 (TReMBrel. 16, Created)
DT	01-MAR-2001 (TReMBrel. 16, Last sequence update)
DT	01-MAR-2001 (TReMBrel. 16, Last annotation update)
DE	CYTCHROME P450.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC	Brassicales; Brassicaceae; Arabidopsis.
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=COLUMBIA;
RX	MEDLINE=98290546; PubMed=9628582;
RA	Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA	Tabata S.;
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RT	Sequence features of the regions of 1,456,315 bp covered by nineteen
RT	physically assigned P1 and TAC clones.";
RL	DNA Res. 5:41-54(1998).
DR	EMBL; AB009048; BAB08653.1; -.
SQ	SEQUENCE 465 AA; 53767 MW; 0C00459C9C866D1F CRC64;

Query Match	18.8%;	Score 481.5;	DB 10;	Length 465;
Best Local Similarity	26.9%;	Pred. No. 3.8e-24;		
Matches 130;	Conservative 110;	Mismatches 197;	Indels 47;	Gaps 13;

[illegible]

OM of: US-09-668-482-2 to: GenEmbl:* out_format : pfs
Date: Nov 6, 2001 3:52 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framet_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09668482/runat_05112001_13353_10884/app-query.fasta_1.1680
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -STAR=1 -MATRIX=bls62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09668482_gcgn1_1_8677 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-668-482-2
Query length: 492

Database: GenEmbl:*

Database sequences: 1344157

Database length: -856060004

Search time (sec): 5019.990000

score_list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation
gb_ov:DRU68234	+ 2563.00	3135.38	2.1e-166	1479	U68234 Danio rerio all-trans-r
gb_ov:AF199462	+ 1816.50	2221.26	1.7e-115	1479	AF199462 Gallus gallus retinoi
gb_pr4:AF005418	+ 1741.00	2127.23	3.0e-110	1743	AF005418 Homo sapiens retinoid
gb_r01:AF115769	+ 1738.00	2123.65	4.7e-110	1726	AF115769 Mus musculus cytochr
gb_r01:MMP450RA	+ 1734.00	2118.59	9.0e-110	1756	Y12657 M.musculus mRNA for P45
gb_ov:AF057566	+ 1646.50	2013.08	6.8e-104	1479	AF057566 Xenopus laevis retin
gb_htg21:AL358613	+ 1279.00	1517.85	2.6e-76	166337	AL358613 Homo sapiens chromo
gb_pr5:AF252297	+ 1034.50	1253.12	1.5e-61	4445	AF252297 Homo sapiens cytochr
gb_ba3:SYCSLLE	+ 693.50	804.57	1.4e-36	113064	D64003 Synecocystis sp. PCC
gb_pat1:AR074363	+ 552.50	672.62	3.1e-29	1608	AR074363 Sequence 1 from paten
gb_p12:ATCYP450R	+ 552.50	672.62	3.1e-29	1608	X87367 A.thaliana mRNA for cyt
gb_p12:AF318211	+ 528.50	643.93	1.2e-27	1494	AF318211 Taxus cuspidata 5-alf
gb_p12:AF279252	+ 515.50	627.07	1.1e-26	1649	AF279252 Vigna radiata cytoch
gb_p11:AC006931	+ 489.00	553.88	1.3e-22	116205	AC006931 Arabidopsis thalian
gb_p14:SLU54770	+ 481.00	586.42	2.0e-24	1395	U54770 Solanum lycopersicum cy
gb_ov:AF185266	+ 460.00	574.74	8.9e-24	322	AF185266 Gallus gallus CYP26 (C
gb_p11:AB008097	+ 455.00	551.46	1.8e-22	1934	AB008097 Arabidopsis thaliana
gb_p12:AF318500	+ 450.00	547.85	2.8e-22	1487	AF318500 Arabidopsis thaliana
gb_p12:AF212991	+ 435.50	528.71	3.2e-21	1718	AF212991 Cucurbita maxima ent-
gb_p11:AB018112	+ 430.00	486.42	7.4e-19	70475	AB018112 Arabidopsis thaliana
gb_p12:AF318501	+ 394.50	478.38	2.1e-18	1740	AF318501 Arabidopsis thaliana
gb_p12:AF326277	+ 384.00	464.97	1.2e-17	1844	AF326277 Hordeum vulgare ent-X
gb_p11:AB020744	+ 377.50	423.35	2.4e-15	62052	AB020744 Arabidopsis thaliana
gb_p14:ZM032579	+ 375.00	454.59	4.4e-17	1724	U32579 Zea mays DWARF3 (dwarf3
gb_ba3:MXTAH	+ 370.50	409.33	8.6e-17	1681	AJ232955 Myxococcus xanthus cy
gb_p13:ATT18B16	+ 363.50	401.97	3.7e-14	96574	AL021687 Arabidopsis thaliana
gb_p12:ATCHRIV50	+ 363.50	395.04	9.1e-14	199382	AL161550 Arabidopsis thalian
gb_ba3:MTCY22D7	+ 359.00	411.98	1.0e-14	31859	Z83866 Mycobacterium tubercul
gb_in2:AF251548	+ 350.50	424.17	2.2e-15	1533	AF251548 Tribolium castaneum c
gb_ov:OMU96077	+ 347.00	421.56	3.0e-15	1802	U96077 Oncorhynchus mykiss cyt
gb_pr5:AF319634	+ 347.00	421.56	3.0e-15	1512	AF280108 Homo sapiens clone 15
gb_ov:GGA250337	+ 346.00	419.57	3.9e-15	1638	AJ250337 Gallus gallus cytochr
gb_r01:AB017785	+ 346.00	417.47	5.1e-15	2041	AB017785 Mus musculus cyp4A10
gb_ba3:SC7E4	+ 344.50	387.95	2.3e-13	36772	AL359214 Streptomyces coelic
gb_ov:RABC450PA	+ 343.50	415.18	6.8e-15	1882	J05034 Rabbit CYP3A6 gene encd
gb_ov:AF251272	+ 343.50	413.19	8.8e-15	2316	AF251272 Oryzias latipes cyto
gb_r01:AF166266	+ 343.00	414.84	7.2e-15	1830	AF166266 Mus musculus Iamoster
gb_p12:AF139532	+ 343.00	414.56	7.4e-15	1883	AF139532 Ligidambar styracifi
gb_pr2:AC007002	+ 342.00	372.07	1.7e-12	140331	AC007002 Homo sapiens BAC cl
gb_pat2:E02279	+ 340.50	411.71	1.1e-14	1843	E02279 Human liver cytochrome

gb_p11:AB037244	+	339.50	412.42	9.8e-15	1506	AB037244 Asparagus officina
gb_p11:AB037245	+	339.50	412.42	9.8e-15	1506	AB037245 Asparagus officina
gb_ov:RABP4503C	+	339.50	410.88	1.2e-14	1768	M19139 Rabbit cytochrome P4
gb_pr10:HDMCYP2C9A	+	339.50	410.47	1.3e-14	1845	M61857 Human cytochrome P45

seq_name: gb_ov:DRU68234

seq_documentation_block:

LOCUS DRU68234 1479 bp mRNA VRT 22-NOV-1996
DEFINITION Danio rerio all-trans-retinoic acid 4-hydroxylase (p450RAI) mRNA,
complete cds.

ACCESSION U68234 GI:1680715

VERSION U68234

KEYWORDS zebrafish.

SOURCE Danio rerio

ORGANISM Danio rerio

REFERENCE 1 (bases 1 to 1479)

AUTHORS White,J.A., Guo,Y.D., Baetz,K., Beckett-Jones,B., Bonasoro,J.,

TITLE Hsu,K.E., Dilworth,F.J., Jones,G. and Petkovich,M.

JOURNAL J. Biol. Chem. 271 (47), 29922-29927 (1996)

MEDLINE 97094702

REFERENCE 2 (bases 1 to 1479)

AUTHORS White,J.A., Guo,Y., Baetz,K., Beckett-Jones,B., Bonasoro,J.,

TITLE E., Dilworth,F.J., Jones,G. and Petkovich,M.

JOURNAL Submitted (28-AUG-1996) Cancer Research Labs, Queen's University,

BTtorell Hall, Kingston, ON K7L 3N6, Canada

FEATURES

source

gene

CDS

1..1479

/gene="p450RAI"

/note="all-trans-retinoic acid 4-hydroxylase"

/codon_start=1

/protein_id="AAC60045.1"

/db_xref="GI:1680716"

/translation="MGLYTLMTFLCTIVLPVLLFLAAYKLMEMIRVDPNCRSPL

PPGTMGLPFIGETLQILLQRRKFLRMKRQKCIYKTHLFGNPTVRVGADNVROILL

GEHLKLVSVQWPAVSVRTILGSDTLNVHGVQHNKKKAIMRAFSRDALHHYIPVIOQEV

KSALQEWLQKDSVLYVPEMKKLMFRIMRILGFEPQIKTDEQELVEAFEEIMKVL

ESLPIDVPEFSGLYRGLRANFIHSKIEENIRKKIQDDNENQKYDALQLLIENSRR

SDEPFSLOAMKEATRELLFGGHETASTATSLVFMFLGNTVEVQKREVEQEKVEMGM

YTPGKGLSMELDQLKTYGCVIKETLRINPVGGFRRVAKTFELNGVOIPKGMNVIV

SICDTHDVADVFPNKEEFQPERFMSKGLDEGSRFNNIIPFGGSRMCVKEFAKVLKKI

FLVELTQHCNMWILNSGPPMTKTGPTIYPVDNLPYTKFTSYVRN"

BASE COUNT 425 a 331 c 379 g 344 t

ORIGIN

alignment_scores:

Quality: 2563.00 Length: 492

Ratio: 5.209 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-668-482-2 x DRU68234 ..

Align seg 1/1 to: DRU68234 from: 1 to: 1479

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17

|||||

1 ATGGGGCTGTACACCTTATGTCACCTTCTGCACCATCTGCTACC 50

|||||

17 oValLeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIlea 34

|||||

51 CGTTTACTCTTCTCTCGCCGGGTGAAGTTGTGGAGATGTTAATGATCC 100
34 rGArgValaspProasnCysArgSerProLeuProGlyThrMetGly 50
101 GACGAGTCGATCCGAACTGCAGAACTCTTACCGCAGGTACCATGGCC 150
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPhe 67
151 TTGCCGTTTCATTGGAGAAACGCTCCAGCTGATCCTCCAGAGAAGAAGTT 200
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
201 TCTGCGCATGAAACGCGCAGAAATACGGGTGCATCTACAAGACGCCACTCT 250
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
251 TCGGGAACCCGACTGTCAGGGTGATGGAGAGCTGATTAATGTGAGGCAGATT 300
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
301 CTGCTGGGCGAACAACAGCTGGTGCTCTTCACTGAGGCCACGATCAGTAG 350
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
351 AACCATCCTGGGCTCTGACACCCCTCTCCAATGTCCATGTGAGTTCAACACA 400
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
401 AAAACAAGAAAAAGGCCATTATGAGGGCGTTCTCTCGAGATGCTCTGAG 450
151 HisTyrIleProValIleGlnGlnLysValLysSerAlaIleGlnGluTr 167
451 CACTACATTCCTCGTGATCCAGACAGAGGTGAAGAGCGCCATACAGAAATG 500
167 pleuGlnLysAspSerCysValleuValTyrProGluMetLysLysLeuM 184
501 GCTGCAAAAAAGACTCTGCGTGGCTTTATCCAGAAATGAAGAAGAACTCA 550
184 etPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
551 TGTTCGGATAGCTATGAGAAATCCTGCTGTGTTTGAACCAAGCAAAATA 600
201 LysThrAspGluGlnGlnLeuValGluAlaPheGluGluMetIleLysAs 217
601 AAGACGAGCAGCAGCAAGACTGGTGGAGCTTTTGAGGAAATGATCAAAAA 650
217 nLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyL 234
651 CTGTGTTCTCTGCAATCGACGCTTCTTTCAGTGTCTGTACAGGGGTT 700
234 euArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArgLys 250
701 TGAGGGCACGCAATTTTCATTCACTCCAAAAATTGAGAAAAACATCAGGAAG 750
251 LysIleGlnAspAspAspAsnGluAsnGluGlnLysTyrLysAspAlaLe 267
751 AAAATTCAAGATGACGACAATGAAAAACGAAACAGAAATACAAAAGACGCCCT 800
267 uGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerLeuG 284
801 TCAGCTGTTGATCGAAGAACAGCAGAAAGATGACGAACCTTTAGTTTGC 850
284 InAlaMetLysGluAlaAlaThrGluLeuLeuPheGlyGlyHisGluThr 300
851 AGCGGATGAAGAAGACAGCTACAGAGCTTCTATTGAGAGTCAATGAACC 900
301 ThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGl 317
901 ACCGCCAGCACTGCAACCTCACTTGTCAATGTTTCTGGGTCTGAACACAGA 950
317 uValValGlnLysValArgGluValGlnGluLysValGluMetGlyM 334
951 AGTGTGCAGAAAGGTCAGAGAGAGGTTCAAGAGAAAGTTGAATATGGCA 1000

334 eTyrThrProGlyLysGlyLeuSerMetGluLeuAspGlnLeuLys 350
1001 TGTATACACCTGGAAGAGGCTTGAGTATGAGCTGTTGACACAGCTGAAG 1050
351 TyrThrGlyCysValIleLysGluThrLeuArgIleAsnProProValPr 367
1051 TACACTGATGTGTGATTAAAGAGACTTTAGAATCAACCCCTCCTGCC 1100
367 oGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnI 384
1101 CGAGAGATTGAGAGTCGCACACTCAAAACCTTGAATTGAATGTTACCAAA 1150
384 leProLysGlyTrpAsnValIleTyrSerIleCysAspThrHisAspVal 400
1151 TTCCTAAGATGGAACGTCATTTTACAGCATCTGTACACCGCAGCATGTG 1200
401 AlaAspValPheProAsnLysGluGluPheGlnProGluArgPheMetSe 417
1201 GCCGACGTCCTTCCAAACAAGAGAGAGTTCCAGCCGAGAGATTCATGAG 1250
417 rLysGlyLeuGluAspGlySerArgPheAsnTyrIleProPheGlyGlyG 434
1251 CAAAGTCTGGAAGGACGGGTCCAGGTTTAACTACATCCCTTCGAGAGAG 1300
434 lYserArgMetCysValGlyLysGluPheAlaLysValLeuLeuLysIle 450
1301 GATCCAGGATGTGTGGCAAAAGAGTTCGCCAAAGTGTACTCAAGATC 1350
451 PheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyPr 467
1351 TTTTATGTTGAGTTAACGACGATTCGAATTGGATTCTCTCAAAACGAGCC 1400
467 oProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuProT 484
1401 CCCGACAATGAAAAACAGCGCCGACTATTACCCAGTGAGACAATCTCCCTA 1450
484 hrLysPheThrSerTyrValArgAsn 492
1451 CCAAAATTCAGTAGTATGTCAAGAAAT 1476
seq_name: gb_ov:AF199462
seq_documentation_block: 1479 bp mRNA VRT 21-NOV-1999
LOCUS AF199462
DEFINITION Gallus gallus retinoic acid degrading enzyme CYP26 mRNA, complete cds.
ACCESSION AF199462
VERSION AF199462.1 GI:6456740
KEYWORDS
SOURCE
ORGANISM chicken.
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 1479)
Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,
Jessel,T.M. and Eichele,G.
Complementary Domains of Retinoic Acid Production and Degradation
in the Early Chick Embryo
Dev. Biol. (2000) In press
JOURNAL
2 (bases 1 to 1479)
Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,
Jessel,T.M. and Eichele,G.
Direct Submission
TITILE Submitted (28-OCT-1999) Max Planck Institute for Experimental
Endocrinology, Feodor-Lynen Strasse 7, Hannover 30625, Germany
JOURNAL
location/Qualifiers
FEATURES
1. 1479
source
CDS
/Organism="Gallus gallus"
/db_xref="taxon:9031"
1. 1479
/note="p450RAI"

```

/codon_start=1
/product="retinoic acid degrading enzyme CYP26"
/protein_id="AAF09250.1"
/db_xref="GI:6456741"
/translation="MGFSALVASALCTELLPLLLFLAAVRLMDLYCASGRDPSCLPLPL
PGTMGLPFFGETLQMLQRRKFLQMKRRKGYFIYKTHLFGRPYVRMGAEENVRHILL
GEHRLVSQWPASVRTILGSGCLSNHNGOHKRRKVIMQAFSRDALQHYVPVIOEEV
SACLAQWLGAGPCLLVYPEVKRLMFRIARILLGFQPRQASPDGEQQLVEAFEEMIRN
LFLSLPIDVPFSGLYRGLRARNIIHAKLEENIRAKMARKEPEGYKDALLQLMHETOGN
GEOLNQELKESATELLFGGHETTASATSLIAFLGLHDVLOKVRKELQDKLGGP
NQEKQLNMEFLQQLKYTCVITKETLRSPVPVGGFRIALKTLELNGYQIPKGMNVIS
ICDFHDVADLFTDKDEFNPDRFMSPSPEDSSRFSEIPEGGLRSCVCGKEFAKVLKIF
TWELARSCDWQLLNGPPTMKTGPVIYPVDNLPAKFIGFSQOI"
BASE COUNT      302 a      438 c      427 g      312 t
ORIGIN

alignment_scores:
  Quality: 1816.50
  Ratio: 4.176
  Percent Similarity: 88.776
  Length: 490
  Gaps: 2
  Percent Identity: 70.408

```

```

alignment_block:
US-09-668-482-2 x AF199462 ..
Align seq 1/1 to: AF199462 from: 1 to: 1479

```

Align seg 1/1 to: AF199462 from: 1 to: 1479

1 MetGlyLeuYrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
 1 ATGGGCTTCCGCTCTGGTCGCGCAGCGCCCTGTGCACCTTCTGTCGCC 50
 17 ovalLeuPheLeuAlaValLysLeuTrpIleMetLeuMetIleA 34
 51 CCTGCTGCTTCCCTGGCCGCGCTCCGGCTCTGGGACCTGTACTGCGCA 100
 34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
 101 GCGGCGCGACCCACAGCTGTCCGCTGCCGTTGCCCCCGGGCACCATGGGG 150
 51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
 151 CTGCCCTTCTTGGGGAGACGCTGCAGATGGTGTCTGCACGGCGGAAATT 200
 67 eleuArgMetLysArgGlnLysTyrglyCysIleTyrglyThrHisLeuP 84
 201 CCTGCAGATGAAGCGCAGAAATACGGCTTCACTACAGACTCACCTCT 250
 84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
 251 TTGGGCGCGCCACCTGTGGGTGATGGCGCTGAGAAATGTGGCGACATC 300
 101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
 301 CTGCTGGGCGAGCACCGCTGTCTCCGTGCAGTGGCCGCGCTCGGTGGC 350
 117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
 351 CACCATCTTGGGCTCGGGCTGCCCTCTCCAACTTCACACGGGCGACACA 400
 134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
 401 AACACCGCAAAAAGGTATCATGCAGGCCCTTCTCCCGGGATGCCCTGCAG 450
 151 HisTyrlleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
 451 CACTACGTGCTGTCAATCAAGAGAGGTGAGTGCCTGCGCTGGCACAGTG 500
 167 pleuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeu 184
 501 GCTGGGCTGCCGCGCCCTGCTGTGTATCCGAGGTGAAGCGCCTCA 550
 184 euPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200

[illegible]

```
seq_name: gb_pr4:AF005418
seq_documentation_block:
LOCUS AF005418 1743 bp mRNA PRI 15-DEC-1997
DEFINITION Homo sapiens retinoic acid hydroxylase mRNA, complete cds.
ACCESSION AF005418
VERSION AF005418.1 GI:2688845
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1743)
AUTHORS White,J.A., Beckett-Jones,B., Guo,Y.D., Dilworth,F.J., Bonasoro,J.,
Jones,G. and Petkovich,M.
TITLE cDNA cloning of human retinoic acid-metabolizing enzyme (hp450RAI)
JOURNAL Identifies a novel family of cytochromes P450
MEDLINE J. Biol. Chem. 272 (30), 18538-18541 (1997)
97373542
REFERENCE
2 (bases 1 to 1743)
AUTHORS White,J.A., Beckett-Jones,B., Guo,Y., Dilworth,F.J., Bonasoro,J.,
Jones,G. and Petkovich,M.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1997) Cancer Research Labs, Queen's University,
Botterell Hall, Rm 355, Kingston, Ont K7L 3N6, Canada
FEATURES
source location/Qualifiers
1..1743
/organism="Homo sapiens"
/db_xref="taxon:9606"
39..1532
/note="hp450RAI; cytochrome P450"
/codon_start=1
/product="retinoic acid hydroxylase"
/protein_id="AAB88846"
/db_xref="GI:2688846"
/translation="MGLPALLASALCTFVLPLLLFLAIAIKLMDLYCVSGRDRSCALPL
PGTGMGFPEFGTLLQVLQRRKFLQMKRRKGYIKTHLFGPTVRVMGADNVRILL
GDDRLSVHMPASVRTILGSGCLSNLHDSHKQKKVIMRAFSALECYVPVITEEV
GSSLEQWLSGGERGLLVPEVKRLMRIALLGCEPOLAGDSEQLYEAFEEEMT
RNLFSLPIDVPEFSGLYRGMKARNLIHARIEONIRAKICGLRASEAGQCKDALQLLIE
HSWGERGLDMQALIKOSSTELLFGGHETTASAATSLITYIGLPHYLQKVRSEILSKG
LICKSNODNKILDMELIEQLKIGCVIKETIRLNPPVPGGFRVALKTFELNGYQIPKGW
NVYISICDTHDVAEIFTNKEEFNPDRFMPHEDASRSEFIPFGGGLRSCVGEFAKI
LKIFVELARHCDWQLNGPPTMKTSPIYYPVDNLPARFTHFHGEI"
BASE COUNT 423 a 442 c 477 g 401 t
ORIGIN
alignment_scores:
Quality: 1741.00 Length: 497
Ratio: 4.087 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.209
alignment_block:
US-09-668-482-2 x AF005418 ..
Align seg 1/1 to: AF005418 from: 1 to: 1743
1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||
:::||||:
39 ATGGGGCTCCGGCGCTGGCCAGTGCCTGTCACCTTCTGCTGCC 88
17 oValLeuLeuPheLeuAlaValIleLeuTyrGluMetLeuMetIleA 34
|:::|||||
:::||||:
89 GCTGCTGCTCTTCTGCTGCTGCATCAAGCTCTGGGACCTGTA 138
34 rGArGValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
||| ::::|
139 GCGGCCGACCGCAGTTGTGCCCTCCCATTGCCCCCGGACTATGGGC 188
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
:::|||||
:::||||:

```

```
189 TTCCCTCTTTGGGAAACCTTGACAGATGTAAGTGCAGCGGAGGAAGTT 238
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
||||:
:::||||:
239 CCTGCAGATGAAGCGCAGAAATACGGCTTCACTCAAGACGCACTCTGT 288
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
||||:
:::||||:
289 TCGGGCGGCCACCGTACGGGTGATGGCGCGGACAATGTGCGGCGCATC 338
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||
:::||||:
339 TTGCTCGAGACGACCGCGCTGTGTCGGTCCACTGCGCAGCGTGGTGG 388
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||
:::||||:
389 CACCATTTCTGGGATCTGGCTGCCTCTCACTGACAGACTCCTCGCAC 438
134 ySAsnLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGln 150
|:::
:::||||:
439 AGCAGCGCAAGAGGTGATATATCGGGCCTTACGCCGAGGACACTCGAA 488
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGlnTr 167
|||||
:::||||:
489 TGCTACGTTGCCGTGATCACCGAGAGAGTGGCCAGCAGCTGAGCAGTG 538
167 pLeuGlnLysAspSerCys.....ValLeuValTyrProGluM 180
|||||
:::||||:
539 GCTG.....AGCTGCGGCGAGCGCGGCTCTGTTCTACCCCGAGG 579
180 eTLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
|:::
:::||||:
580 TGAAGCGGCTCATGTTCCGAATCGCATGCGCATCTACTGGGCTGCGAA 629
197 ProGlnGlnIleLysThrAsp.....GluGlnGluLeuValGluAl 210
||| ||:::
630 CCC...CACTGCGGCGGCGAGGGAAGTCCGAGCAGCAGCTTGAGAGGC 676
210 apPheGlnGluMetIleLysAsnLeuPheSerLeuProIleAspValProp 227
|||||
:::||||:
677 CTTCGAGGAATGACCGCGCAATCTCTCTCGCTGCCCATCGAGCTGCCCT 726
227 heSerGlyLeuTyrArgGlyLeuArgAlaArgAsnPheIleHisSerLys 243
|||||
:::||||:
727 TCAGCGGCTGTACCGGGCGATGAAGCGCGGAACCTCATTCACGCGCGC 776
244 IleGlnGluAsnIleArgLysIleGlnAsp...AspAspAsnGluAs 259
|||||
:::||||:
777 ATCGAGCAGAACATTCGCGCCAGATCTGCGGCTCGGGCATCCGAGGC 826
259 nGluGlnLysTyrLysAspAlaLeuGlnLeuIleGlnAsnSerArgA 276
|||
827 GGGCCAGGGCTGCAAGAGCGGCTGCAGCTGTGATCGAGCACTCGTGGG 876
276 rGSerAspGluProPheSerLeuGlnAlaMetLysGluAlaAlaThrGlu 292
::|
877 AGAGGGGAGAGCGGCTGACATGCAGGCACCTAAGCAATCTCAACCGAA 926
293 LeuLeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerLeuVa 309
|||||
:::||||:
927 CTCTCTTTGGAGACACGAAACACGCGCAGTGCAGCCACATCTGTGAT 976
309 lMetPheLeuGlyLeuAsnThrGluValValGlnLysValArgGluGluV 326
:|
977 CACTTACCTGGGGCTGTACCAACATATGTCTCCAGAAAGTGCAGAGAGGC 1026
326 aGlnGlnLysValGluMetGlyMetTyrThrProGlyLysGlyLeuSer 342
:::
1027 TGAAGAGTAAGGTTTACTTTGCAAGAGCAATCAAGACAACAGTTGGAC 1076
343 MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGluTh 359
|||||
:::||||:
1077 ATGGAATTTTGAACAACACTTAATACATCGGGTGTATTAAAGAGAGAC 1126

```


359 rleuArgIleasnProValProGlyGlyPheArgValAlaLeuLysT 376
|||||:|||||
1127 CCTCGACTGAATCCCCAGTTCACAGAGGGTTCCGGCTTGGCTCGAAGA 1176
376 hrpHegluLeuAsnGlyTyrGlnIleProLysGlyTrrPAsnValIleTyr 392
|||||
1177 CTTTGATTAATGATGATACAGATTCACCAAGGCTGAATGTATCTAC 1226
393 SerIleCysAspThrHisAspValAlaAspValPheProAsnLysGluI 409
|||||
1227 AGTATCTGTGATCTCATGATGTGGCAGAGATCTTCACCAAGAAGA 1276
409 upHegInProGluArgPheMetSerLysGlyLeuGluAspGlySerArgP 426
|||||
1277 ATTTAATCCTGACCGATTCATGCTGCTCACCACAGAGATGCATCCAGGT 1326
426 heAsnTyrIleProPheGlyGlyGlySerArgMetCysValGlyLysGlu 442
|||||
1327 TCAGCTTCATTCATTTGGAGGAGGCTTAGAGAGCTGTAGGCAAGAA 1376
443 PheAlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCy 459
|||||
1377 TTTGCAAAATTTCTTCAAAATATTTACAGTGGAGCTGGCCAGGCATTG 1426
459 sAsnTrrPileuSerAsnGlyProProThrMetLysThrGlyProThri 476
||:|||||
1427 TGACGTGCAGCTTCTAAATGACCTCCTACATGAAGAAACAGTCCACCG 1476
476 leTyrProValAspAsnLeuProThrLysPheThrSerTyr 489
:::|||||
1477 TGTATCTGTGACATCTCCCTGCAAGATTCAACCATTTTC 1517
seq_name: gb_rol:AF115769

seq_documentation_block: 1726 bp mRNA ROD 03-MAR-1999
LOCUS AF115769 1726 bp mRNA ROD 03-MAR-1999
DEFINITION Mus musculus cytochrome P450 retinoic acid metabolizing enzyme
P450RA mRNA, complete cds.
ACCESSION AF115769
VERSION AF115769.1 GI:4324987
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1726)
AUTHORS Paine,C.T., Paine,M.L. and Snead,M.L.
TITLE Identification of amelo-genin- and tuftelin-interacting proteins
using the yeast two-hybrid system
JOURNAL Connect. Tissue Res. 38, 257-267 (1998)
REFERENCE 2 (bases 1 to 1726)
AUTHORS Paine,C.T., Paine,M.L. and Snead,M.L.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1998) CMB, University of Southern California,
2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
FEATURES
source location/Qualifiers
1. 1726
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="tooth"
/dev_stage="1 day old"
27. 1520
/note="tuftelin-interacting protein 24 TIP-24"
/codon_start=1
/product="cytochrome P450 retinoic acid metabolizing
enzyme P450RA"
/protein_id="AAD17217.1"
/db_xref="GI:4324987"
/translation="MGIPLALATALCTFVLPILLFLAALKLMDLYCVSSRDRSCALPL
PGTMGFPEFGETLQMLQRRKFLQMKRRKYGYIKTHLFGPTVRVMGADNVRILL
GEHLRVSVMWPAASVRTILGAGLSNLHDSHQRKRVIMQAFSREALQCYLVIAEEV
SSCLEQWLSCGERGLLVYPEVKRLMFRIAMRILLGCEPDAGGEDEQQLVEAFEEMT

BASE COUNT 405 a 428 c 476 g 417 t
ORIGIN
RNLFSLPIDVPFSGLYRGVKARNLIHARIEENIRAKIRRLQATEPDGCKDALQLLIE
HSWGERGLDMQALKQSSSTELLFGHETASATSLITYGLYPHVLQKVRREIKSKG
LLCKSNQDNKIDMETLEQLKYIGCVIKETLRNLPVPGGFRVALKTFELNGYQIPKGM
NVIVSICDTHVDADIFTNKEEFNPDRIYHPEDASRFSFIPGGGLRSCVGEFAKI
LLKIFTVELARHCDWQLLNGPPTMTKTSPTVYPVDNLPAFTYRQDGI"
alignment_scores: Length: 496
Quality: 1738.00 Gaps: 4
Ratio: 4.080
Percent Similarity: 85.887 Percent Identity: 67.944
alignment_block:
US-09-668-482-2 x AF115769 ..
Align seg 1/1 to: AF115769 from: 1 to: 1726
1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||
27 ATGGGCTCCCGCGGCTGCTGGCCACTGCGCTCTGCACCTTCGTGCTGCC 76
17 ovalLeuLeuPheLeuAlaAlaValLysLeuTrrPgluMetLeuMetIleA 34
||:|||||
77 GCTGTGCTCTTCTGCGGCGGCTCAAGCTCTGGGACCTGTACTGTGTGA 126
34 rGArgValAspProAsnCysArgSerProLeuProProGlyTrrMetGly 50
||:|||||
127 GCAGCCGCGATCGCAGCTGCGCCCTCCCTGCCCCCGGTACCATGGGC 176
51 leuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
:::|||||
177 TTCCCATTTCTTGGGGAACATTGCAGATGCTGCTTCAGCGGAGGAGTT 226
67 eleuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
|||||
227 TCTGCAGATGAAGCGCAGGAATAACGGCTTCATCTACAAGACGATCTGT 276
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|||||
277 TTGGCGCGCCACGCGTGGCGGTGATGGCGCGGATTAATGTGGCGGCATC 326
101 leuLeuGlyGluHisLysLeuValSerValGlnTrrProAlaSerValAr 117
|||||
327 TTGCTGGAGAGACACCGGTGGTGTGCGTGCACGTGCCCGCGTGGTGG 376
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||
377 CACCATCTTGGCGCGTGGCTGCCCTCTCCAACTGCACGATTCCTCGCACA 426
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
|||||
427 AGCAGCGAAAGAGGTGATATGACAGGCCCTTCAGCCGAGGACTCCAG 476
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTrr 167
|||||
477 TGCTACGTGCTGTCGATCGCTGAGGAAGTCAGACAGTTGTCGAGCAGTG 526
167 pleuGlnLysAspSerCys.....ValLeuValTyrProGluM 180
|||||
527 GCTA.....AGCTGCGGCGAGCGCGGCTCCTGCTCTACCCCGAGG 567
180 etLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
:::|||||
568 TGAAGCGCTCATGTTCGCGATCGCATGCGCATCTGCTGGCGTGGAG 617
197 ProGluGlnIle.....LysThrAspGluGlnGluLeuValGluAlaPh 211
|||
618 CCGGGTCCAGCGGCGGCGGCGGAGACGACACAGCTCTGGAAGGCTTT 667
211 eGluGluMetIleLysAsnLeuPheSerLeuProIleAspValProPheS 228
|||||

668	CGAGGAGATGACCCCGCAATCTCTCTCTCTCTCCCATGACGTGCCCTTTA	717
228	ergIlyleuTyrrargglyleuArgAlaargAsnpheIleHisSerLysIle	244
718		767
245	gluGlulAsnIleargglyLysIleglInasp..AspaspasngluAsnGI	260
768	GAGGAGAACATTCGCGCCCAAGATCCGCCGGCTTCAGGCTACAGAGCCGGA	817
260	uglnLsTyrrLysAspAlaleuGlInleuIleugluAsnSerArgArGs	277
818	TGGGGGTTCGAAGGAGCGCCTGCAGCTCCTGATTCAGCACTCGTGGGAGA	867
277	erAspgluProPheSerleuGlInAlametLysgluAlaAlaThrgluleu	293
868	GGGAGAGAGAGCGTGGATATGCAGGCACTAAACATCGTCACACAGAGCTC	917
294	leuPheglglyHisgluThrThrAlaserThrAlaThrSerleuValme	310
918	CTCTTGTGTGTCATGAACACTACAGCCAGTGTCTCGACATCCTGATCAC	967
310	tpheleuglyleuAsnThrgluValValglInLysValArggluIluValG	327
968	TTACCTAGAGACTCTACCCACATGTCTCTCCAGAAAGTTGCAGAAAGATAA	1017
327	InglulysValgluMetgluMetTyrrThrProgllyLysglyleuSermet	343
1018	AGAGCAAGGGCTTACTTTGCAAGAGCAATCAAGACACAAGTTAGACATG	1067
344	gluleuLeuAspglnleuLysTyrrThrglyCysValIleLysgluThrle	360
1068	GAAACTTTGGGAACAGCTTAATATCATTTGGGTGTTCATTAAAGAGACCTT	1117
360	uArgIleasnProProValProglglyPheargValAlaleuLysThrp	377
1118	GCGATTGAATCTCCGGTCCAGAGGGTTTCGGGTGCTCTGAAGACTT	1167
377	hegluleuAsnglyTyrrglInleProLysglyTrpasnValIleYrSer	393
1168	TTGAGCTGAATGATACAGATCCCAAGGGCTGGAATGTTATTTCAGT	1217
394	IleCysAspThrHisaspValAlaaspValPheProAsnLysgluIuph	410
1218	ATCTGTGACACCCACGATGTGGCAGATATCTTCACTAAACAAGAGGAATT	1267
410	eglProgluArgPheMetSerLysglyLeuGluaSpglySerArgPhea	427
1268	TAAATCCGACCGCTTATAGTGCCTCATCCAGAGGATGCTCCGGTTCA	1317
427	snTyrrIleProPheglglyLysArgMetCysValglYlysgluPhe	443
1318	GCTTCATTCATTGGAGAGAGGCTTCGGAAGCTGTGTAGGCAAGATT	1367
444	AlaLysValleuLeuLysIlePheleuValgluleuThrghInHisCysAs	460
1368	GCAAAATAATCTTCTTAAGATATTTACAGTGGAGCTGGCTAGGCACGTGA	1417
460	nTrpIleleuSerAsnglyProProThrMetLysThrghlyProThrIleT	477
1418	TTGGCAGCTTCTAATATGAGACTCTCAATGAAGACACAGCCCACTGTGT	1467
477	yrProValAspAsnleuProThrLysPheThrSerTyrr	489
1468	ACCCTGTGACAATCTCCCTGCACAAGATTCCACCTACTTTC	1505

seq_name: gjb_rol:MMP450RA

seq_documentation_block:		
LOCUS	MMP450RA 1756 bp	08-JAN-1998
DEFINITION	M.musculus mRNA for P450RA protein.	
ACCESSION	Y12657	
VERSION	Y12657.1 GI:2765213	

KEYWORDS	P450RA gene.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 1756) Fujii,H., Sato,T., Kaneko,S., Gotoh,O., Fujii-Kuriyama,Y., Osawa,K., Kato,S. and Hamada,H.
TITLE	Metabolic inactivation of retinoic acid by a novel P450 differentially expressed in developing mouse embryos
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1756)
AUTHORS	Hamada,H.
TITLE	Direct Submission
JOURNAL	Submitted (17-APR-1997) H. Hamada, Osaka University, Institute for Mol. & Cell. Biol., 1-3 Yamada-Oka, Suita, Osaka, 565, JAPAN
FEATURES	Location/Qualifiers
source	1..1756 /organism="Mus musculus" /strain="C3H" /db_xref="taxon:10090" /cell_line="p19" 57..1550 /gene="P450RA" 57..1550
CDS	

BASE COUNT	ORIGIN
408 a	434 c
494 g	420 t

```
alignment_scores:  
    Quality: 1734.00  
        Ratio: 4.070  
Percent Similarity: 85.887  
  
                Length: 496  
                    Gaps: 4  
Percent Identity: 67.742
```

```
alignment_block:
US-09-668-482-2 x MMP450RA . .
```

Align seg 1/1 to: MMP450RA from: 1 to: 1756

```

1 MetGlyLeuTYrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
   |||||  ::|||:|||||  |||||:|||||
57 ATGGGGCTCCGGCGCTGTGGCCAGTGGCTGTGCACCTTGtGTGCC 106
   |||||  ::|||:|||||  |||||:|||||
17 oValLeuLeuPheLeuAlaIleValLysLeuTrpGluMetLeuMetIleA 34
   ||::|||  |||||:|||||  |||||:|||||  ::|||
107 GCTGTGCTCTTCTCCTGGCGGCGCTCAAGCTGTGGACCTGTACTGTGA 156
   |||||  ::|||:|||||  |||||:|||||  |||||
34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetCly 50
   |||  ::|||  |||||:|||||  |||||:|||||
157 GCAGCGCGATCGCAGCTGCGCCCTCCCTTGCCCCCGGTACCATGGGC 206
   |||  ::|||  |||||:|||||  |||||:|||||
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
   ::|||:|||||  |||||:|||||  |||||:|||||
207 TTCCCATCTTTGGGAAACATTGCAGATGTGTCTTACGCGAGGAAGTT 256
   |||||  ::|||:|||||  |||||:|||||
67 eLeuArgMetLysArgGlnLysTrpGlyCysIleTyrrLysThrHisLeuP 84
   |||||:|||||  |||||:|||||  |||||:|||||

```

[illegible]

377 heglIeuAsnglyTYrglnIleProLysGlyTRpaSnValIleTYrSer 393
|||||
1198 TTGAGCTGAATGATGATACCAGATCCCAAGGCTGGAATGTTATTACAGT 1247
394 IleCysasprThriSaspValAlaaspValPheProAsnLysGluGluPhe 410
|||||
1248 ATCTGTGACACCCACGATGTGGCAGATATCTTCACCTAACAAAGAGAATT 1297
410 eglInProgluArgPheMetSerLysGlyLeuGluAspGlySerArgPheA 427
|:::|
1298 TAATCCCGACCGCTTTATAGTGCCCTCATCCAGAGAGATGCTCCCGGTTCA 1347
427 snTYrIleProPheGlyGlyGlySerArgMetCysValGlyLysGluPhe 443
:::|
1348 GCTTCATTCATTGGAGGAGGCCCTTCGAGAGCTGTGTAGGCAAGAGATT 1397
444 AlAlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAs 460
|||||
1398 GCAAAATTTCTTTAAGATATTTACAGTGGAGCTGGCTAGGCACTGTGA 1447
460 nTYrIleLeuSerAsnGlyProProThrMetLysThrGlyProThrIleT 477
::|
1448 TTGGCAGCTCTTAATGACCTCCTACAAATGAAGACAAGCCCACTGTGT 1497
477 yrrProValAspAsnLeuProThrLysPheThrSerTYr 489
|||||
1498 ACCCTGTGGACAATCTCCCTGCAAGATTCACCTACTTC 1535

seq_name: qb_ov:AF057566

seq_documentation_block:

DEFINITION *Xenopus laevis* retinoic acid converting enzyme mRNA, complete cds.

VERSION AF057566.1 GI:3283171

SOURCE African clawed frog.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,

Xenopodinae; *Xenopus*.

AUTHORS Hollemann, T., Chen, Y., Grunz, H. and Pieler, T.

acid signalling
7361 7373 (1008)

MEDLINE 99077803

AUTHORS
Chen, J. L., Grunz, H., Panitz, F., Pieler, T. and Hollemann, J.

JOURNAL

Submitted (05-APR-1998) Developmental Biochemistry, University of
Göttingen, Number 32, Coeditorin Lower Saxony 37073.

Germany

source 1. .1479

```

/db_xref="taxon:8355"
1_1479
cds

```

```
/product="relinoic acid converting enzyme"
/protein_id="AAC25158.1"
```

```

/ab_xref=GL:32831/2
/translacion="MDLYTLI

```

PPGIMGLPEFGEILQMWLQKKRF LQVKNSSQIGNIKNHLEQSF I VAV ICHENV KXZL
GEKLI SVHWPASVRTILGAGCLSNLHDNEHKYTKKVIAQAFSREALANYVPQMEEEV

KCSVNLMEQSGPCVLIPAIKNNMFRIAMKLLDSCDYADNRYDAZLNZLNDNRG
LESLPIDVPESGLYRGHARNLIHAOIEENIKEKLÖREPEDEHCKDALÖLLIDYSRRNG

EPINLQALNESAI EELFGHG I IASARISUIS EELINDV EENAKNEB I EEEB I
EKKEL SIEVL OOLKYTSCVIKETLR LSPVAGGF RVALKTFVLNGYQIPKGNVIYS

VEICRNCDEWELNGSPAMTTSPIICPVNDLPAKFKPFSSSI"

```

BASE COUNT      350 a      422 c      384 g      323 t
ORIGIN

alignment_scores:
    Quality: 1646.50      Length: 494
    Ratio: 3.892      Gaps: 4
    Percent Similarity: 85.628      Percent Identity: 63.158

alignment_block:
US-09-668-482-2 x AF057566 ..

Align seg 1/1 to: AF057566 from: 1 to: 1479

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuP 17
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 ATGATCTGTACACTCTGCTCACACGCGCTGTGTGCACCCCTGGCAGTGC 50

17 oValLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
  |:::| ||||| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
51 ATGTGCTGCTCTCTCACCGCTGCACCAACTGTGGGAAGTCTACTGCTGA 100

34 rGArgValAspProAsnCysArgSerProLeuProGlyThrMetGly 50
  ||||| ||| ||| ::||| ||||| ||||| ||||| ||||| |||||
101 GGCGCAAGATGCCCGCTGCGCTAACCCACTGCCCCGGCACCATTGGGA 150

51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysP 67
  ||||| ||::| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 CTGCCCTTCTTCGAGAGACTCTGCAATGTGTCTACAGAGCGCAGGTT 200

67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeu 84
  ||||| ::::| ||||| ::::| ||||| ||||| ||||| ||||| |||||
201 CCTCCAAGTGAAGCGAAGCCAAATACGGTTCGATCTACAAGACGCATCTGT 250

84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
  ||||| ::::| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 TCGGCAGCCCCACGGGTGCGCTCACGGGCGCGAGAACGTTGCAGAGATC 300

101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
  ||:::| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 CTGATGGGGGAACACAAAGCTGTGTGCGTGACGTGGCGGCGCTGTGTGCG 350

117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
  ||||| ||||| ::::| ||||| ::::| ||||| ::::| ||||| ::::|
351 CACCAATCCTAAGGGCGCGCTGTGTCTCCAACTGCACGACACACGACACA 400

134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
  ||| ||||| ::::| ||| ::||| ||||| ||||| ||||| ||||| |||||
401 AGTACACCAAGAAGGTGATCGCGCAAGCCTTCTCCCGGGAAGCCCTCGCT 450

151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGluTr 167
  ::|||::||| ::::| ||||| ::::| ||||| ::::| |||
451 AATTACGTGCCGAGATGGAAGAGAGGTGAGATGTCTCCGTGAACCTGTG 500

167 pleuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeuM 184
  ||||| ::::| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 GCTGCAGAGCGGCCCCGTGTGTCTGTATATCCCGGCATCAACGAATGA 550

184 etPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 TGTTCGCATTGCCATGAGGCTCTGCTGCGGTGTGATCCCAACGCAATG 600

201 LysThrAspGluGlnGlu...LeuValGluAlaPheGluGluMetIleLy 216
  ::::| ::::| ||| ||::| ||||| ||||| ||||| ||||| |||||
601 GACCGCAACAGAGAAACGCTGCTCGAAGCCTTCGAGGAATGAGCCG 650

216 sAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg 233
  ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 AAATCTTTTCTTGGCTATGTATGTGCGCTTTAGCGGCTCTACCGGG 700

233 LysLeuArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArg 249
  ||||| ||||| ||||| ::||| ::::| ::::| ::::| ::::| ::::|
701 GTCTGGCGGGCCAGAACTAATTCATGCCCCCAATTGAGGAAAAACATCAAG 750

```

```

250 LysLysIleGlnAspAspAsnGluAsnGluGlnLysTyrLysAspAl 266
    ::::::::::: ::::: ::::: :::::
751 GAAAAGCTGCAA.....AGAGAACCAAGATGAACATTGCAAGATGC 791

266 aLeuGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerL 283
    ||||| ||||| ::::: ||||| ::::: ||||| :::::
792 CTTGCAGCTGCTGATTGACTATAGCCGACGAAATGGCGACCATCAACT 841

283 euGlnAlaMetLysGluAlaAlaThrGluLeuLeuPheGlyGlnIsglu 299
    ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
842 TGCAGGCATTGAAAGAGTCCGCAACTGAGCTCTGTTTGAGGCCACCGA 891

300 ThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnTh 316
    ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
892 ACCACGGCAAGTGCCTCCACAGTCCCTCACCCTCTTTCTTCGCCCTTCACAA 941

316 rGluValValGlnLysValArgGluGluValGlnGlnLysValGluMetG 333
    ::::: ||||| ::::: ||||| ::::: ||||| :::::
942 AGACGTTCTGGAAAAGGTCCGCCAAGAGCTTGAACACAG.....G 982

333 lYMetTyrThr.....ProGlyLysGlyLeuSerMetGluLeu 345
    ||::: ::::: ||| ||||| ::::: ||||| :::::
983 GTCTCTGTCTCACTAAACCCGAGGAAAAGAAGCACTGACGATCGAAGTA 1032

346 LeuAspGlnLeuLysTyrThrGlyCysValIleLysGluThrLeuArgI 362
    ||::: ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1033 CTTCAACAGATTAAAGTACTAGCTGTGTCATTAAAGAGACTCTGCGGCT 1082

362 eAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu 379
    ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1083 CAGCCACACAGTGGCCGGTGGATTTCGGTGGCTCTCAAGACGTTGTAT 1132

379 euAsnGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSerIleCys 395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1133 TAAATGGCTATCAGATTCCCAAGGGCTGGAATGTATATACAGCATTTGCT 1182

396 AspThrHisAspValAlaAspValPheProAsnLysGluGluPheGlnPr 412
    ||||| ||||| ||||| ::::: ||||| ::::: ||||| :::::
1183 GACACCCACGCGTAGGCCGATCTGTTCCCGACACGCAACAATTCAACCC 1232

412 OGluArgPheMetSerLysGlyLeuGluAspGlySerArgPheAsnTyrI 429
    ||::: ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1233 TGACCGCTTTCTAACGCCACTTCCCAGAGATTCTCAAGGTTTGCTTCA 1282

429 lProPheGlyGlyGlySerArgMetCysValGlyLysGluPheAlaLys 445
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1283 TCCCTTTGGAAGGTGGAGTGAGGTGCTGCATCGGCAAGAATTGTGCTAAA 1332

446 ValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAsnTyrI 462
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1333 ATCTTCTCAAGGTGTTGCTGTGGAATGTGTGCTAATTGCGACTGGGA 1382

462 eLeuSerAsnGlyProProThrMetLysThrGlyProThrIleTyrProV 479
    ||||| ||||| ||::: ||||| ||::: ||||| ||||| |||||
1383 GCTCTTGAACGAGATCCCCCGCATGACAACCTCCCCGATTATCTGTCTG 1432

479 aLasPasnLeuProThrLysPheThrSerTyr 489
    ||||| ||||| ::::: ||||| ::::: ||||| :::::
1433 TGGACAACCTACCCGCTAAATTCAAAACCTTC 1464

seq_name: gb_htg21:AL358613

seq_documentation_block:
LOCUS AL358613 166337 bp DNA HTG 14-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-348J12, *** SEQUENCING IN
PROGRESS ***, 14 unordered pieces.
ACCESSION AL358613
VERSION AL358613.8 GI:13660841
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

REFERENCE      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS        Plumb,B.
TITLE          Direct Submission
JOURNAL        Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
                CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
                requests: clonerequest@sanger.ac.uk
                On Apr 17, 2001 this sequence version replaced gi:13625010.
COMMENT        ----- Genome Center
                Center: Sanger Centre
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquerry@sanger.ac.uk
                ----- Project Information
                Center project name: BA348J12
                ----- Summary Statistics
                Assembly program: XGAP4; version 4.5
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Consensus quality: 161818 bases at least Q40
                Consensus quality: 163255 bases at least Q30
                Consensus quality: 163957 bases at least Q20
                Insert size: 165037; sum-of-contigs
                Insert size: 176845; 5.5% error; agarose-1p
                Quality coverage: 5.93x in Q20 bases; sum-of-contigs Quality
                coverage: 5.81x in Q20 bases; agarose-1p
                -----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 14 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *
                * 1 32409: contig of 32409 bp in length
                * 32410 32509: gap of 100 bp
                * 32510 52913: contig of 20404 bp in length
                * 52914 53013: gap of 100 bp
                * 53014 55817: contig of 2804 bp in length
                * 55818 55917: gap of 100 bp
                * 55918 74512: contig of 18595 bp in length
                * 74513 74612: gap of 100 bp
                * 74613 83833: contig of 9221 bp in length
                * 83834 83933: gap of 100 bp
                * 83934 87001: contig of 3068 bp in length
                * 87002 87101: gap of 100 bp
                * 87102 97488: contig of 10387 bp in length
                * 97489 97588: gap of 100 bp
                * 97589 100583: contig of 2995 bp in length
                * 100584 100683: gap of 100 bp
                * 100684 105295: contig of 4612 bp in length
                * 105296 105395: gap of 100 bp
                * 105396 129702: contig of 24307 bp in length
                * 129703 129802: gap of 100 bp
                * 129803 132462: contig of 2660 bp in length
                * 132463 132562: gap of 100 bp
                * 132563 137099: contig of 4537 bp in length
                * 137100 137199: gap of 100 bp
                * 137200 158309: contig of 21110 bp in length
                * 158310 158409: gap of 100 bp
                * 158410 166337: contig of 7928 bp in length.
                *
                FEATURES
                source
                1. 166337
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="10"
                /clone="RP11-348J12"
                /clone_1b="RPC1-11.2"
                1. 32409
                /note="assembly_fragment:02593
                fragment_chain:1
                clone_end:T7

```

[illegible]

63 63
55164 GCTGAAGTCGGGGTAGCGCCCGGAGGACATGCTATGCGGCTAGGAG 55213
63 63
55214 CAGGCTGGCGGAGCGCGGCTCCCGCGCCCTCATGCCCACCTTC 55263
64 ArgArgLysPheLeuArgMetLysArg 72
55264 TCCTCTCCGCTTCTCTCCACAGCGAGGAAGTCTCTCGGAGATGAAGCGC 55313
73 GlnLysTyrGlyCysIleTyrLysThrHisLeuPheGlyAsnProThrVal 89
55314 AGGAATAACGGCTTCATCTACAGACGATCTGTTCGGGCGGCCACCGT 55363
89 LArgValMetGlyAlaAspAsnValArgGlnIleLeuGlyGluHisL 106
55364 ACGGCTGATGGCGCGGACACAATGCGGCGCATCTGTCTCGAGAGACACC 55413
106 LysLeuValSerValGlnTyrProAlaSerValArgThrIleLeuGlySer 122
55414 GCGTGGTGTGCGTCCACTGGCCAGCGTGGTGGCACCATTCTGGGATCT 55463
123 AspThrLeuSerAsnValHisGlyValGlnHisLysAsnLysLysLys.. 138
55464 GCGTGCCTCTCTAACCCTGACGACTCTCTCGACAAGCAGCGCAAGAAGT 55513
138 138
55514 GGGGCGAGGAGCGGCGCTGACAGGAGGGGACCCCATTTATGAGCG 55563
138 138
55564 GAATTCGGCGCTGATGATGCTAGCGCGGCTAGCAGCTTGAGGTGGGCT 55613
138 138
55614 AGGACCTCTGCGCAGCTCCAGGTTAGCTATACCCAGCTCGGAGAGTGCCA 55663
138 138
55664 TGTGTCTGGCAGGACTGGGGTGTCTGGAAGGGGAGCGGGGTAGACGAGA 55713
139 Ala.II 140
55714 GGGCGGATGAGGCTTTTAACCTGTCCCTCTCGGACTCAGGTGAT 55763
140 eMetLArgAlaPheSerArgAspAlaLeuGlu..... 150
55764 TATCGGGCTTCAGCGCGGAGCACTCGAATGCTACGTGCGCGGTGATCA 55813
150 150
55814 CCGANN 55863
150 150
55864 NNN 55913
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluT 167
55914 NNNNTACGTGCGGTGATCAACGAGGAAGTGGGAGAGCCTGGAGCAGT 55963
167 rPheUGlnLysAspSerCys.....ValLeuValTyrProGlu 179
55964 GGCTG.....AGCTGGCGGAGAGCGCGGCTCTCTGTCTACCCCGAG 56004
180 MetLysLysLeuMetPheArgIleAlaMetArgIleLeuGlyPheG1 196
56005 GTGAAGCGCTCATGTTCGAATCGCCATGCGCATCTACTGGGCTGCGA 56054
196 uProGlnGlnIleLysThrAsp.....GluGlnGluLeuValGlu 210

56055 ACCC...CAACTGGCGGGCGGACGGGGACTCCGAGACAGCAGCTTGTGAGG 56101
210 LArgGluGluMetIleLysAsnLeuPheSerLeuProIleAspValPro 226
56102 CCTCGAGGAATGACCCCGCAATCTTCTCTGCTGCCCATCGACGTGCC 56151
227 PheSerGlyLeuTyrArg..... 232
56152 TTCAGCGGGCTGTACCGGGTTAAGGGCGCAACGGGCTGCGGACTAGGGG 56201
233Gly. 233
56202 CGCGGACCTGGGCGTCTGCTCACCAGCGCGGCTCTCTGCGCTCAGGGC 56251
234 LeuArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArgLy 250
56252 ATGAAGCGCGGAACTCATTTACGCGCGCATCGAGCAGAACATTCGCGC 56301
250 sLysIleGlnAsp...AspAspAsnGluAsnGluGlnLysTyrLysAsp 266
56302 CAAGATCTGGGCTGCGGCGATCCGAGGCGGGCCAGGCTGCAAGACG 56351
266 LLeuGlnLeuLeuIleGluAsnSer..... 274
56352 CGCTGACGCTGTGATCGAGCAGCTCGTGGAGAGGGAGAGCGGCTGGAC 56401
274 274
56402 ATGACAGTGAATGACAGCTTCAGACAGGCACTGCGGAGTTGTCCCT 56451
274 274
56452 GCGTTTCCAAGCGTGTCTCTGGGCGCCCAAGCGCGCTGGGGCC 56501
274 274
56502 AGCTTCTGAGTGGCGCGCGGCTCAGACTACAGCTATGGAATCCCGAA 56551
275ArgArgSerAspGlu..... 279
56552 GGAAGGCTGAGACACCGGTACAGAGAGCTCGGGAAGGGGCTGCGCGGA 56601
279 279
56602 ACTGGAGCATCCCTAGCCTTTCAGGTTTCAAAAGGAAGTTGGAAT 56651
279 279
56652 TGCAAAATGTATAAAGAACCCTTGCATTTTAATAAAACTAAGACTTT 56701
280ProPheSerLeuG1 284
56702 AACTCAGAGTTCCGGTAGAGCGGGTGTACTCGCCTTACTG:CTCCA 56750
284 n..... 284
56751 GCTGAACATAAGGAGCGTTGCATTTTGTTAAGATATGCTTCCCTTGA 56800
284 284
56801 CTTTCTGTACGAAACATTTAGCCCTTCTAGTCTTCCCTCCAGAACTCT 56850
284 284
56851 CAGTTGATTTGAGTAATCCTTGTGCAAAACCGCAGGAGACTTGTGAG 56900
285AlaMetLysGluAlaIleThrGluL 293
56901 AATGTGGTCTCACTATTTCTTAGGACTTAAAGCAATCTTCAACCGAAC 56950
293 eLeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerLeuVal 309
|||||

```

56951 TCCTCTTTGGAGGACACGAAACACGGCCAGTGCAGCCACATCTCTGATC 57000
310 MetPheLeuGlyLeuAsnThrGluValAlaGlnLysValArgGluGluVala 326
      :::::::::::::::::::::
57001 ACTTAAGTGGGGCTCTACCCACATGTTCTTCAGAAAGTCCGAGAAGAGCT 57050
326 IGIInGluLysVal...GluMetGlyMet..... 334
      :::::::::::::::::::::
57101 AAATTGAGCTGCTCCCTAGCCAACTTCCGAATAAGTCAGTGTGCTGCTT 57150
334 ..... 334
57151 CATGAGTATTTTGAAGAAGTCAAGGCCAGGGCTGCGTGCGCCAGTGGGC 57200
334 ..... 334
57201 AGAATTAGCTTTGTGAATAAACAAGATGAGTCTTGAGCTGTGAGCCAC 57250
334 ..... 334
57251 TTGGGACAGGGTTGACCTGTATTCCTACCCCTCCTCCACCTTTGCTGA 57300
334 ..... 334
57301 ACCGTGAGATTCTCAGATAGGTTCCACTTCTTGAATTGGTGTCCAA 57350
334 ..... 334
57351 GGGCATACATAGTGTGAATAGCTGATTAGTGTGGGTGGTGTGAGTG 57400
334 ..... 334
57401 TGGGTGGGGTTTTTTTGGGGGAACCTGTGCTCTTAAGTGTGAGCAG 57450
335 TyrThrPro.....GlyLys.Gly. 340
      :::::::::::::::::::::
57451 CATTCTCCTGGGATGTGAATAGATAGTGTGGATTTGGCAGGCAAAATGCC 57500
340 ..... 340
57501 ATTAGCTGCTGTTTCCCCCTCCAGTCTCTCCCATCATGGGCCCTTGG 57550
340 ..... 340
57551 GTTTAGTCTCCACTTAAGCCCTGTTACGTCCTGCTGGGCTGATTTTATTG 57600
340 ..... 340
57601 GAGCACAATAAATACTGTTACACTCTGTATGACTGTTTGATAGGGTTAC 57650
341 .....LeuSerMetGluLeuLeuAspGln 348
      :::::::::::::::::::::
57651 TTTCGAAGACCAATCAAGACAACAAGTTGGACATGGAATTTTGAAACAA 57700
349 LeuLysTyrThrGlyCysValIleLysGluThrLeuArgIleAsnProPr 365
      :::::::::::::::::::::
57701 CTTAATAATACATCGGGGTGTGTTATTAAAGAGACCTTCGACTGAATCCCCC 57750
365 oValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsn... 380
      :::::::::::::::::::::
57751 AGTTCAGGAGGTTTCGGGTGTGCTCTGAAGACTTTGAATTAAATGTAA 57800
380 ..... 380
57801 GTTAAATTTCTCTTCTTCTCCCTTTTGTTGTGTGTTAAAACTCCTCTT 57850
380 ..... 380
57851 TCTTCCCTATGCTGTGGCTGCACATTCCTATGCTTTTGATTAATGTTCTGC 57900

```

```

380 ..... 380
57901 CCTATATGAGATATGTTTCAGCAACCTGGATCCACTCTCTTTCTAC 57950
380 ..... 380
57951 CTCTCCCTTGCTTTAGCTCATTAATTTCCCCCAAGATATCAGTACTGC 58000
380 ..... 380
58001 TTTTGTGTGTGAAGTTAAATTCAGTTTGTACGCCCTTGGGTTTCAT 58050
381 ..... Gly. TyrGlnIleProLysGlyTrpAsnValIleTyrse 393
58051 AATCTTTTGCAGGATACCAGATTCCTCAAGGCTGGAATGTATCTACAG 58100
393 rIleCysaspThrHisaspValAlaaspValpheProAsnLysGluLup 410
58101 TATCTGTGATCTCATGATGTGGCAGAGATCTTCACCAACAAGAAGAT 58150
410 heGlnProGluArgPheMetSerLysGlyLeuGluAspGlySerArgphe 426
58151 TTAATCCTGCACCGATTTCATGCTGCTCACCAGAGAGATGCATCCAGTTTC 58200
427 AsnTyrIleProPheGlyGlySerArgMetCysValGlyLysGluPh 443
58201 AGCTTCATTCATTGGAGAGAGGCTTAGGAGCTGTGTAGGCAAGAATT 58250
443 eAlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysA 460
58251 TGCAAAAATCTCTTCAAAATATTTACAGTGGAGCTGGCCAGGCAATTGTG 58300
460 snTriPleuSerAsnGlyProProThrMetLysThrGlyProThrIle 476
58301 ACTGGCAGCTTCTAAATGGACCTCCTACAAAGAAACAGTCCCAACGTG 58350
477 TyrProValaspAsnLeuProThrLysPheThrSerTyr 489
58351 TATCCTGTGGACAACTCTCCCTGCAAGATTCACCACTTTTC 58389
seq_name: gb_pr5:AF252297

seq_documentation_block:
LOCUS AF252297 4445 bp mRNA PRI 14-JUN-2000
DEFINITION Homo sapiens cytochrome P450 retinoid metabolizing protein
P450RAI-2 mRNA, complete cds.
ACCESSION AF252297
VERSION AF252297.1 GI:8515440
KEYWORDS
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE
AUTHORS White,J.A., Ramshaw,H., Taimi,M., Stangle,W., Zhang,A.,
Everingham,S., Creighton,S., Tam,S.P., Jones,G. and Petkovich,M.
IDENTIFICATION of the human cytochrome P450, P450RAI-2, which is
predominantly expressed in the adult cerebellum and is responsible
for all-trans-retinoic acid metabolism
Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6403-6408 (2000)
20300913
2 (bases 1 to 4445)
White,J.A., Ramshaw,H., Taimi,M., Stangle,W., Zhang,A.,
Everingham,S., Creighton,S., Tam,S.-P., Jones,G. and Petkovich,M.
Direct Submission
Submitted (04-APR-2000) Cytochroma Inc., 116 Barrie Street,
Bioscience Complex, Suite 2424, Kingston, ONT K7L 3N6, Canada
Location/Qualifiers
1. .4445
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .1539

```


[illegible]

```

seq_name: gb_pat1:AR074363

seq_documentation_block:
LOCUS      AR074363      1608 bp      DNA      PAT      28-AUG-2000
DEFINITION Sequence 1 from patent US 5952545.
ACCESSION  AR074363
VERSION    AR074363.1  GI:10001118
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
            Unclassified.

REFERENCE  1 (bases 1 to 1608)
AUTHORS   Koncz,C., Mathur,J., Szekeres,M. and Altmann,T.
TITLE     Nucleic acid molecules encoding cytochrome P450-type proteins
          involved in the brassinosteroid synthesis in plants
          Patent: US 5952545-A 1 14-SEP-1999;
          Location/Qualifiers
FEATURES   Location/Qualifiers
            source          1..1608
                           /organism="unknown"

BASE COUNT 382 a      374 c      374 g      478 t

ORIGIN

```

alignment_scores:		
Quality:	552.50	Length: 485
Ratio:	1.811	Gaps: 14
Percent Similarity:	62.087	Percent Identity: 29.897

```

alignment_block:
US-09-668-482-2 x AR074363      ..

Align seg 1/1   to: AR074363   from: 1   to: 1608

```

```

19 LeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleArgAr 35
   |||||:::||||:::||||:::||||:::||||:::||||:::||||:::
66 CTCCTCCTCCTCTCTTCATCGCGCGCGGCTTCTCCTCTACTCCGCCG 115
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

35 gValAspProAsnGlyArgSerProLeuProGlyThrMetGlyLeuP 52
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
116 TACACGTTACCGT...CGGATGGGTCTGCCCTCCGGGAAGCCTTGTCCTC 162
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

52 roPheIleGlyGluThrLeuGlnLeuIle.....LeuGln 63
   ||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
163 CTCGTATAGAGAGACTTTTCAGCTGATCGGAGCTTACAAACAGAGAAC 212
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

64 ArgArgLysPheLeuArgMetLysArgGlnLysThrGlyCysIleTyrL 80
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
213 CCTGAGCCTTTCATCGACGAGAGAGTAGCCCGGTACGGTTCCGTTTTCAT 262
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

80 sThrHisLeuPheGlyAsnProThrValArgValMetGlyAlaAspAsnV 97
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
263 GACGCATCTTTTGGGTGAACCGACGATTTTCTCAGCTGACCCGGAACGA 312
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

97 AlaArgGlnIleLeuLeuGlyGlnHisLysLeuValSerValGlnTrpPro 113
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
313 ACCGGTTGTCTTCAGAACGAAGGGAAGCTTTTGAGTGTTCTTATCCT 362
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

114 AlaSerValArgThrIleLeuGlySerAspThrLeuSerAsnValHisGl 130
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
363 GCTTTCATTTGTAAACCTTTGGGGAAACACTCTCTGCTTCTTATGAAGG 412
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

130 yValGlnHisLysAsnLysLysLysAlaIleMetArgAlaPheSerArgA 147
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
413 TTTCTTGCATAAA...CGTATGCACCTCTCACCATGAGCTTTGCTAATT 459
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

147 sPalaLeu.....GluHisTyrIleProValIleGlnGlnGluValLys 161
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
460 CTTCAATCATTTAAAGACCATCTCATGCTTGATATGACCGGTAGTCGG 509
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

162 SerAlaIleGlnGluTrpLeuGlnLysAspSerCysValLeuValTyrPr 178
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
510 TTTAATCTTGATTTCTTGG.....TCTTCTCGTGTCTCCTCATGGA 550
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

178 oGluMetLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyP 195
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
551 AGAAGCCAAAAGATAACGTTTGAGCTAACGGTGAAGCAGTTGATGAGCT 600
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

195 heGluPro.....GluGlnIleLysThrAspGluGlnGluLeu 207
   ||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
601 TTGATCCAGGGGAATGAGTGAAGATTTTAAGGAAGAGTATCTTCTTGTG 650
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

208 ValGluAlaPheGluGluMetIleLysAsnLeuPheSerLeuProIleAs 224
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
651 ATCGAAGGCTTC.....TTCTCTCTTCTCCTCCTCCC 679
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

224 pValProPheSerGlyLeuTyrArg...GlyLeuArgAlaArgAsnPheI 240
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
680 TCTC...TTCTCCACCACTTACCGCAAGCCATCCAAAGCGCGAGGAAGG 726
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

240 leHisSerLysIleGluGluAsnIleArgLysLysIleGlnAspAspAsp 256
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
727 TGGCGGAGCGGTTGACGGTGTGTGATGAAGAAAGG...AGGAGAGAGAG 773
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

257 AsnGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGluAs 273
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
774 GAAGAAGGAGCGGAGAGAAGAAGATATGCTTGCGGCTTGCTT..... 818
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

273 nSerArgArgSerAspGluProPheSerLeuGlnAlaMetLysGluAlaA 290
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
819 .....GCGGCGGATGATGATTTTCCGATGAAGAGATTGTGACTTCT 861
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

290  lathrgluLeuLeuPheGlyGlyHisGluThrThrAlaSerThrAlaThr 306
    ::  |||||  ::|||::||| |||||:::  ||
862  TGGTGGCTTACTGTTCGCCGGTTATGAACAACCTCCACGATCATGACT 911
307  SerLeuValMetPheLeuGlyLeuAsnThrGluValAlaGlnLysValAr 323
    ||  |||||  ::  :::::  :::::
912  CTGCCGCTCAAAATTCTCACCGAGACTCCTTAGCTCTTGCTCAACTCAA 961
323  ggluGluValGlnGluLysValGluMetGlyMetYThrProGlyLysG 340
    :|||  ::|||:::  ::||  ::  :::
962  GGAAGAG...CATGAAAAGATTAG...GCAATGAAGAGTGCATTCGTATA 1005
340  lyleuSerMetGluLeuLeuAspGlnLeuLysTyThrGlyCysValIle 356
    ::|||::  ::  :::::  ::||  |||||::
1006  GTCTTGAATGAGAGTGATTACAAAGTCAATGCCATTCCACACAATGTGTGTT 1055
357  LysGluThrLeuArgIleAsnProProValProGlyGlyPheArgValAl 373
    ::||| |||||::  ::  ||  |||||  ||
1056  AATGAGACGCTACGAGTGCTAACATCATCGCGGTGTTTCAGACGTGC 1105
373  aleuLysThrPheGluLeuAsnGlyTyrglnIleProLysGlyTrpAsnV 390
    |:::  ||:::  |||:::  ||||| |||||:::
1106  AATGACGAGATGTTGAGATCAAAAGTTATAAAATTCCAAAAGGTGGAAG 1155
390  alileTySerIleCysAspThrHisAspValAlaAspValPheProAsn 406
    ||::  ||::  ::||  ::  ||  :::
1156  TATTCTCATCGTTTAGAGCGGTTTCATTAGACCCCAACCACTTCAAGAT 1205
407  LysGluGluPheGlnProGluArgPheMetSerLysGlyLeuGluAspGl 423
    ::  ||::||  ||:::  ||:::  ||:::
1206  GCTCGCACTTCAACCCCTTGAGATGGCAGAGCAACTCGTTAACGACAGG 1255
423  ySerArgPheAsnTyriIleProPheGlyGlySerArgMetCysValG 440
    |  ::  ||||| |||||  ||:::  ||
1256  CCCTTCTAATGTGTTACACACCGTTTGTGGAGGGCCAAAGGCTATGTCCCG 1305
440  lylYsgluPheAlaLysValLeuLysIlePheLeuValGluLeuThr 456
    ||  ||:::  ||:::  ||  :::::  ||||  ::||:::
1306  GTTACGAGCTGGCTAGGTTGCACCTCTGTGTTCTTCACCGCCTAGTG 1355
457  GlnHisCysAsnTrpIleLeuSerAsnGlyProProThrMetLysThrGl 473
    ::|||::  ||::  ||:::  ::
1356  ACAGGCTTCAGTTGGGT.....CCTGCAGAGCAAGACAA 1390
473  yProThrIleTyrrProValAspAsnLeuProThrLysPheThrSerTyrv 490
    :::::  ||::  ::  :::::  ::||
1391  GCTGTTTCTTCTTCCAACTACAGAAGCAGAAACGGTACCCTGATCTTCG 1440
490  alArg 491
    ||::
1441  TGAAG 1445

seq_name: gb_pl2:ATCYP450R

seq_documentation_block:
LOCUS  ATCYP450R 1608 bp mRNA PLN 13-MAY-1996
DEFINITION  A.thaliana mRNA for cytochrome P450.
ACCESSION  X87367
VERSION  X87367.1 GI:853718
KEYWORDS  Cytochrome P450.
SOURCE  thale cress.
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 1608)
AUTHORS  Szekeres,M., Nemeth,K., Koncz-Kalman,Z., Mathur,J., Kauschmann,A.,
Altman,T., Redei,G.P., Nagy,F., Schell,J. and Koncz,C.
JOURNAL  Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
MEDLINE  controlling cell elongation and de-etiolation in Arabidopsis
CELL  Cell 85 (2), 171-182 (1996)
REFERENCE  2 (bases 1 to 1608)

```

AUTHORS Szekeres, M.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-1995) M. Szekeres, Inst. of Plant Biology,
Biological Research Center, Hungarian Academy of Sciences, PO Box
521, 6701 Szeged, HUNGARY

FEATURES
source
1. .1608
/organism="Arabidopsis thaliana"
/cultivar="columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone_lib="lambda gt10"
/clone="C204"
/map="14.3 cm"
48. .1466
/gene="CPD"
48. .1466
/gene="CPD"
/function="loss of gene function results in dwarf
phenotype"
/note="cytochrome P450"
/codon_start=1
/product="CYP90 protein"
/protein_id="CAA60793.1"
/db_xref="GI:853719"
/db_xref="SWISS-PROT:Q42569"
/translation="MAFTAFLLLLSSIAAGFLLLRTRYRRMGLPGLIGET
FOLIGAYKTENPEPFIDERVARYGSVFMTLGEPTIFSADPETNRYVLQNEKLFEC
SYSPASICNLGKHSLLMKSLHMRHSLTMSFANSIIKDMLMDIDRLVRFNLSW
SSRVLLMEBAKKITFELTVKQLSFDPGEWSESLRKEYLVVIEGFSPLPLFSTYR
KAIQARRKVAEALTVVVMKRREEEEGAERKMDLAALLAADGFSDEEIVFLVAL
VAGYETSTIMTLAVKFLFETPLALAQLKEEHEKIRAKKSDSYSLEMSDYKSMPTQC
VWNETLRVANIIGGVERRAMTDVEIKYKIPKGMKVESSEFRVHLDPNHFKDARTFN
WRWQNSVTTGPSNVFTPEGGPRLCPGYELARVALSVFLHRLVTGFSWVPAEQDKLV
FEPTRTQKRYPFVKKRDFEAT"

BASE COUNT 382 a 374 c 374 g 478 t
ORIGIN

alignment_scores:
Quality: 552.50 Length: 485
Ratio: 1.811 Gaps: 14
Percent Similarity: 62.887 Percent Identity: 29.897

alignment_block:
US-09-668-482-2 x ATCYP450R ..

Align seg 1/1 to: ATCYP450R from: 1 to: 1608

```

19 leuLeupheleuAlaValLysleuTrpGluMetleuMetileuArgar 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 CTCCTCCTCCTCTCTCCATCGCCGCCGCTTCCCTCCTCCTACTCCGCCG 115
35 gValAspProAsnCysArgSerProleuProProGlyThrMetGlyleuP 52
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
116 TACACGTTACCGT...CGGATGGGTCTGCCTCCGGGAAGCCTTGCTCC 162
52 ropheleleGluThrleuGlnleuIle.....leuGln 63
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
163 CTCGTGATGAGAGAGACTTTTCAGCTGATCGAGCTTACAAAACAGAGAAC 212
64 ArgArgGlySphleuArgMetLysArgGlnLysTyrgLysIleTyrlly 80
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
213 CCTGAGCCTTTCATCGACGAGAGAGTAGCCCGGTACGGTTCCGGTTTCAT 262
80 sThrHisleuPheGlyAsnProThrValArgValMetGlyAlaAspAsnV 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 GACGCACTTTTGTGGTGAACGACGATTTTCTCAGCTGACCCGGAACGA 312
97 alArgGlnIleleuLeuGlyGluHisLysleuValSerValGlnTrpPro 113
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
313 ACCGGTTTGTCTTCAGAACGAAGGGAAGCTTTTGTAGTGTCTTATCCT 362

```


gene	CDS	BASE COUNT	ORIGIN
gene	CDS	437 a 414 C 384 g 414 t	
<pre> /cultivar="2937" /db_xref="taxon:3916" 1..1649 /gene="C1pCYP" 91..1515 /gene="C1pCYP" /codon_start=1 /product="cytochrome P450" /protein_id="AAF89209.1" /db_xref="GI:9587211" /translation="MVSPLTLLFLFAASAAIFLHRAFSRRKFERLPPGSGLPFIGET LQLISAVKSSNPEPEFMDERVRRYGSIFMTMVEGEPTVFSADEPLNRFILONEGKLDDC SYPGSISNLLGKHSLLMLMGALHKRMHSLTMSFANSIIKDHLHHIDRLIGLNDTW SDRVTLNDQAKKITFELTVKQLMSFDEMTESLRKEYLVLEGFTLPLPSTTYR RAIKARTVAEALTIVVRORREERYNOGKEKSDMLGALLASGDHFSDDOIVDFLLALL VAGYEFTSTIMTLAVKFLTETPLAOLKEEHDQIRASDGPALFETWIDYKSMVFTQH VVNETLRYANIIIGIFERRATTDIDIKYTIIPKGWKEASFRVHLNPEYYKDAFTFNP WRMSQSSSEANPANVYTPGCGPRLCGYELARVVLVSFLHRIIVRFSWVPAEEDKLL VFEPTRTQKRYPPIIVKRRNANHV" </pre>			

```
alignment_scores:
```

Quality:	515.50	Length:	480
Ratio:	1.713	Gaps:	15
Percent Similarity:	62.708	Percent Identity:	29.375

```
alignment_block:
US-09-668-482-2 x AF279252 ..
```

Align seg 1/1 to: AF279252 from: 1 to: 1649

[illegible]

173	svalleuValTyrProGluMetLysLysLeuMetPheArgIleAlaMetA	190
174	:::: ::: ::: ::::	
580	.GTTACTCTCATGACCACGCAAAAAAGATAACGTTTGAAGCTAACACATGA	628
190	rgIleLeuLeuGlyPheGluProGluGlnIleLysThrAspGluGlnIu	206
629	AGCAATGATGAGCTTTGACCCCTGATGAATGG.....ACTGAGAGT	669
207	LeuValGluAlaPheGluGluMetIleLysAsnLeuPheSerLeuProI	223
670	TTAAGGAAGAGTATGTTCTTGTCATTTGAAGGCTTCTCACTCTCCCTTT	719
223	easPValProPheSerGlyLeuTyr..ArgGlyLeuArgAlaArgAsn	239
720	ACCTCTC...TCTCCACCACATACCGGAGAGCCATCAAGCAAGAACAA	766
239	heIleHisSerLysIleGluGluAsnIleArgLysIleGlnAspAsp	255
767	AGGTGCAGAGGCACCTGACGTTGGTAGTCAGGCAGAGG..AGAGAAGAG	813
256	AspAsnGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuLeuIleG	272
814	TATAACCAAGGGAAGAGAAAAAGAGTAGACATGCTAGGGGCACCTGTG..	861
272	uasSerArgArgSerAspGluProPheSerLeuGlnAlaMetLysGluA	289
862GCCCTCCGGTGACCACCTTTCCGACGACCAATAATAGTCGATT	901
289	IaAlaThrGluLeuLeuPheGlyGlyHisGluThrThrAlaSerThrAla	305
902	TCTGTGGCTTGTGCTCGTCGCCGGTTACGAGACCACCTCTACCATATG	951
306	ThrSerLeuValMetPheLeuGlyLeuAsnThrGluValValGlnLysVa	322
952	ACTCTTGCAGTCAAGTTCCTCACTGAGACTCCTCTGGCTCTGGCACAGCT	1001
322	IArgGluGluValGlnGluLysValGluMetGlyMetTyrThrProGlyL	339
1002	CAAGGAAGAG...CATGACCAAATCAGA..GCAAGAGGTGACCCCGGGG	1045
339	ysGlyLeuSerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCysVal	355
1046	CTCCACTGGATGACAGATGACAGATTACAGTCAATGGTTTAAACCAACATGTT	1095
356	IleLysGluThrLeuArgIleAsnProProValProGlyGlyPheArgVa	372
1096	GTAATGAGACCTTGAGGGTGCAACATAATTGGTGGGATATTTAGAAG	1145
372	IAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrP	389
1146	AGCAACGACAGACATCGACATAAAGGTTACACTATTCCTAAGGATGGA	1195
389	snValIleTyrSerIleCysAspThrHisAspValAlaAspValPhePro	405
1196	AGGCTCTTGCAATTCATTTCGTGCTGATCAATCTGAATCCCTGAATATTACAA	1245
406	AsnLysGluGluPheGlnProGluArgPheMetSerLysGlyLeuGluAs	422
1246	GATGCACGCACTTCAACCCCTGGAGATGGCAGAGTAATCATCGGAAGC	1295
422	pGlySerArgPheAsn..TyrIleProPheGlyGlySerArgMetC	438
1296	AGCAAAACCTTCAAAATGTTTATACACCTTTTGGTGGAGGCCACGGTTGT	1345
438	ysValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeuValGlu	454
1346	GCCCTGGTTATGAGCTAGCCAGAGTTTACTTTCGTCTTCCTTCAACCGC	1395
455	LeuThrGlnHisCysAsnTyrIleLeuSerAsnGlyPro.....	467
1396	ATTGTACACCCGCTTCAAGCTGGGTTCTCTGTAAGGAAGATAAGTTGGTGTT	1445
468	...ProThrMetLysThrGlyProThrIleTyrProVal	479

||||| :||| |||||
1446 TTTCACACACAGCAGCAGAGAGG...TACCCTATT 1482

seq_name: gb_pl1:AC006931

seq_documentation_block:

LOCUS AC006931 116205 bp DNA PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 231 of 255 of the
complete sequence. Sequence from clones F14N22, F7D19.

ACCESSION AC006931 AE002093

VERSION AC006931.5 GI:6598641

KEYWORDS HTG.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 116205)

AUTHORS

Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D.,
Fuji, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E.,
Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M.,
Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., Vanaken, S.E., Umayam, L.,
Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H.,
Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D.,
Mierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and
Venter, J.C.

Sequence and analysis of chromosome 2 of the plant Arabidopsis

thaliana

JOURNAL Nature 402 (6763), 761-768 (1999)

MEDLINE 20083487

PUBMED 10617197

REFERENCE 2 (bases 1 to 116205)

AUTHORS Lin, X.

TITLE Direct Submission

Submitted (09-MAR-2000) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA

On Dec 17, 1999 this sequence version replaced gi:4512656.

The sequence and annotation of chromosome 2 were merged from those

of the individual clones on this chromosome after removing

overlaps. For detailed information, please see the TIGR web site

(http://www.tigr.org/tdb/at.html).

COMMENT

JOURNAL

TITLE

AUTHORS

REFERENCE

PUBMED

MEDLINE

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota

Viridiplantae

Embryophyta

Tracheophyta

Spermatophyta

Magnoliophyta

eudicotyledons

core eudicots

Rosidae

eurosids II

Brassicales

Brassicaceae

Arabidopsis

thaliana

thale cress

HTG

AC006931

AE002093

AC006931.5

GI:6598641

PLN

05-APR-2000

231 of 255

of the

complete

sequence

from clones

F14N22,

F7D19.

Sequence

from clones

F14N22,

F7D19.

Department of Energy and the US Department of Agriculture.
Address all correspondence to: at@tigr.org.
Location/Qualifiers
1. 116205

FEATURES

source

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="II"

/complement(<1. .2131)

/note="Sequence from clone F14N22"

<1226. .>2344

/gene="At2g42650"

<1226. .>2344

/gene="At2g42650"

/note="F14N22.8; predicted by genscan"

1226. .2344

/gene="At2g42650"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAF18664.1"

/translation="MSRVSPKTVDDAVKALYKGNKSRTEKPOLLEDEGEFYLVAL
KIPQNFNTNARIPPLPPLINTEDSPCLIIDRPEGLTEPAKKNIKSENIPI
TKVVKLSKLSKSDYSGESKRSKLCDSYDGMFPRVPLPMLPKLIGKKEFSKTPVAID
LKHNMWKEQIEKACGAAMFMRGSCSAIKYAKLSMESDDIVENTATLNGVVDVLP
RWKYIRSLHLKLSLPLVQTPYQIKIDPLVEEVKNGEGLAKSDVDDSSKSY
KTKKNGKIHEVRVYMDSNVSETLGDDEFDRSVGEDEVADDLNAGSGKRRKKKMSSSKS
AVSGKPDIVKSKNGQSKKLKIDDESQSGEFKAKTKRR"

/complement(2132. .>116205)

/note="Sequence from clone F7D19"

3450. .3489,3570. .3646,3732. .>3954))

/gene="At2g42660"

/complement(<2761. .>3954)

/gene="At2g42660"

/note="F7D19.34; predicted by genscan"

3450. .3489,3570. .3646,3732. .3954))

/gene="At2g42660"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD21740.1"

/db_xref="GI:4512686"

/translation="MAGRSSPGKCSWTLDLNRETEVEKNEEESVEDEDQITNNVRQ
YRSNMPRLRTPLDLHSFVRAYORLGPDRAPKLVLEMNKLGSIAHVSHLQMY
RSKLEPSSRPGEFAMSGQSYLMDIDSRICPHSDLRHAYNSKYTPVLDQAVY
TNLGNFLMRPSSWFSGLCRNDRDSTENKTLPLEIRKKNNEKRVDEEVSSVKRLKS
MSGGCIQLPEFGNCRQKPTDDINTMLSLSLSTSSSEP"

/complement(3681. .3724

/rpt_family="(GA)n"

4087. .4204

/rpt_family="(TA)n"

5756. .5806

/rpt_family="(GA)n"

/complement(join(<7700. .7818,7906. .8112,8195. .8348,

8844. .8989,9294. .>9396))

/gene="At2g42670"

/complement(<7700. .>9396)

/gene="At2g42670"

/note="F7D19.33; predicted by genscan"

8844. .8989,9294. .7818,7906. .8112,8195. .8348,

/gene="At2g42670"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD21739.1"

/db_xref="GI:4512685"

/translation="MPYEQRLYNTCKASFSDDPITDALAEKVRNVEKIKPSDVGI
EODQOLARSRSGLPLNERNNGSNOSPRAIKYILHLHEDCSFSIGIFCMPPSSMIPLHNHG
MTVLKSLVYGSMAVSVYDWLEPQTEPEPSQEARPAKLVKDTMTAQSPTTLYPKS
GNIHCFKAITHCAILDILAPPYSEHHRHCTYFRKSRREDLPGLEVDGEVVTDTW
LLEFPDPDFVIRIRIPYRGPIRT"

/complement(join(7700. .7818,7906. .8112,8195. .8348,

8844. .8989,9294. .>9396))

/gene="At2g42670"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD21739.1"

/db_xref="GI:4512685"

/translation="MPYEQRLYNTCKASFSDDPITDALAEKVRNVEKIKPSDVGI
EODQOLARSRSGLPLNERNNGSNOSPRAIKYILHLHEDCSFSIGIFCMPPSSMIPLHNHG
MTVLKSLVYGSMAVSVYDWLEPQTEPEPSQEARPAKLVKDTMTAQSPTTLYPKS
GNIHCFKAITHCAILDILAPPYSEHHRHCTYFRKSRREDLPGLEVDGEVVTDTW
LLEFPDPDFVIRIRIPYRGPIRT"

/complement(join(7700. .7818,7906. .8112,8195. .8348,

8844. .8989,9294. .>9396))

/gene="At2g42670"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD21739.1"

/db_xref="GI:4512685"

/translation="MPYEQRLYNTCKASFSDDPITDALAEKVRNVEKIKPSDVGI
EODQOLARSRSGLPLNERNNGSNOSPRAIKYILHLHEDCSFSIGIFCMPPSSMIPLHNHG
MTVLKSLVYGSMAVSVYDWLEPQTEPEPSQEARPAKLVKDTMTAQSPTTLYPKS
GNIHCFKAITHCAILDILAPPYSEHHRHCTYFRKSRREDLPGLEVDGEVVTDTW
LLEFPDPDFVIRIRIPYRGPIRT"

/complement(join(7700. .7818,7906. .8112,8195. .8348,

8844. .8989,9294. .>9396))

/gene="At2g42670"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD21739.1"

/db_xref="GI:4512685"

/translation="MPYEQRLYNTCKASFSDDPITDALAEKVRNVEKIKPSDVGI
EODQOLARSRSGLPLNERNNGSNOSPRAIKYILHLHEDCSFSIGIFCMPPSSMIPLHNHG
MTVLKSLVYGSMAVSVYDWLEPQTEPEPSQEARPAKLVKDTMTAQSPTTLYPKS
GNIHCFKAITHCAILDILAPPYSEHHRHCTYFRKSRREDLPGLEVDGEVVTDTW
LLEFPDPDFVIRIRIPYRGPIRT"

This work was supported by the National Science Foundation,


```

repeat_region 10075..10153
                /rpt_family="(GAA)n"
mRNA           join(10610..10809,11165..11248,11522..11604,11693..11994)
                /gene="At2g42680"
gene           10610..11994
                /gene="At2g42680"
                /note="F7D19.32"
CDS            join(10668..10809,11165..11248,11522..11604,11693..11812)
                /gene="At2g42680"
                /note="unknown protein"
                /codon_start=1
                /protein_id="AAD21738.1"
                /db_xref="GI:4512684"
                /translation="MAGIGPITQDWEPPVIRKKPANAAAKRDEKTVNARRSGADIEF
                VRKFNAGTNKAASGTSINTKMLDDDTENDTHERVPTLKKAIMOARTDKKLTQSOLA
                QIINEKPOVIOEYESGKAIPNOQILSKLERALGAKLRGKR"
mRNA           complement(join(11994..12273,12362..>13378))
                /gene="At2g42690"
                /complement(11994..>13378)
gene           /gene="At2g42690"
                /note="F7D19.31"
CDS            complement(join(12052..12273,12362..13378))
                /gene="At2g42690"
                /codon_start=1
                /product="putative lipase"
                /protein_id="AAD21737.1"
                /db_xref="GI:4512683"
                /translation="MATTTISWELLGSKNWDTLIDPLDQSLRELILRCDFEQATYD
                AFVNDQNSKYCGASRYKSSSFEDKVMLENASDYEYVNFYATARVSLPEGLLOSQR
                DSWDRSNWFGYIAVTSDESKALGRREIYIALRGTSRNEYVWNLGARPSADPLIH
                GPEQDSSGVVEGTFPDSDEDEEGCKWMLGWLITYTSNPESKFTKLSLSOLLAKI
                KELLEKYKDEKPSIVLTGHSLGATEAVYLAADIAENGSSDPVPTAIVFGCPQGNKE
                FRDEVASHKNLKIHLVARNITDLTRYPGGLIGYVDIGINVYIDTKKSPFLSDSRNPGD
                WHNLOAMLHVAVAGWNGKKEFKLMVKRSIALVNVKSCFEFLAECLVPGSWWEKKNKGLI
                KNEDGEVVLAPVEEPEPVEF"
mRNA           join(<14160..14207,14422..14565,14779..14890,15255..15434
                15513..15664,15832..17364,17446..17514,17602..17786)
                /gene="At2g42700"
                /gene="At2g42700"
                /note="F7D19.30"
CDS            join(14160..14207,14422..14565,14779..14890,15255..15434,
                15513..15664,15832..17364,17446..17514,17602..17730)
                /gene="At2g42700"
                /note="unknown protein"
                /codon_start=1
                /protein_id="AAD21736.1"
                /db_xref="GI:4512682"
                /translation="MALIDVAISCLNSIREIEDVKDAIYIDAGCTESFQFGAFL
                FLEIGARAVCSLENMTSLDVAADWNSKSDCAKRIVIMTSLRLINDAHRMLRCLSTHEG
                PDAYREYETLLVQDYNEHTKSDKISKDGVCSFSSALESTLMEPIESEYNDISSGA
                QGSVAEASLSRQHEDSLSGFLPIISTGSMDDTDVPPGATLTAHFLYQLAKMELKLE
                IFSLGDQSKNVGKILTDMSVVYDVARRRKSGALLVDRITDILTPCCHGDSLFDRIFS
                SLPRAEFRSSQAQLKQGVPIINRPSLDVQVPLGELLNEEFSKIRDSGLPEGTEAFLRG
                WDSYTSAPQNVGLFNECDKSTTNMTSLINGSLVATECFGTPLYEAMIDRKTGDSY
                LVKKMLQELALRENISYVNRARPGYATPELOAMIKALSOSSSLKKNKGIIOGAAT
                AAALDESQSAKWDTFSSAEMMLNVSAGDPSOGIAAOISDLINKSAVAEIOAKKNEKPD
                SSSRGLISFRDALLITVGYILAGENEPTSGSGGPFSSWQEHFLKEAIVDAVLENPSA
                GNKPLNGLTLEELTEGRNLRLKSEETKEIIPSDQDLIDALDDDPWGKMGDEEEEVDS
                KADESYDDMQLKLDLRVDSLFRFLHKLSSLRTRNLPUREGSLASESSFPGEPSGNK

```

67376	TTGGTTCTCTCATATGGGTCCTTCTTAAGCATAGCTGCACAACATCTCCAGCAC	67425
18	lleuLeupheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleArgA	35
67426	ACCTTTCTTCTTCCGC....AAGAAACATCACCGATTTCATAACAAGA	67469
35	rgValAspProAsnCysArgSerProLeuProProGlyThrMetGlyLeu	51
67470	AGATTGAGAAAG....AAAAAGAGCTTCTCTCGAGAAATGGGCTT	67513
52	ProPheIleGlyGluThrLeuLeuIleuIleLeuGlnArgArgLys.....	66
67514	CCATGGATTTGGTGAACAACATGATTTCTACAAAGCTCAGAAAAAGCAACAG	67563
67PheLeuArgMetLysArgGlnLysTyrgLysIleT	79
67564	AGTTTTTGAGGATTTTGTAAATCCAAGCATCATCAACATGGGAACATCT	67613
79	yrLysThrHisLeuPheGlyAsnProThrValArgValMetGlyAlaAsp	95
67614	TCAAGACAGAAATCATGTGGATCACCACAACATAGTAGTCATGGAGCTGA	67663
96	AsnValArgGlnIleLeuLeuGlyGlnHisLysLeuValSerValGlnTr	112
67664	GCCAAATAGGCTCATATATATCCAACGAGTTTAGCTTGGTGGAGCTCGTG	67713
112	prOAlaSerValArgThrIleLeuGlySerAspThrLeuSerAsnValH	129
67714	GCCTTCACTCTTCTGTTCAACTAATGGGCAATGCACTGCATCATGGCTAAC	67763
129	IsGlyValGlnHisLysAsnLysLysLysAlaIleMetArgAlaPheSer	145
67764	AAGCGGAAAGCATCGGGTCTCAGAGGAGATTGTAGCTAATAGTCTTAGC	67813
146	ArgAspAlaLeuGlnHisTyrlleProValIleGlnGlnValLysSe	162
67814	TACATTGGGTTAGAGTCCCTTAATACCTAAGCTCTGTGATACCGTTAAGTT	67863
162	ralaIleGln..GluTrpLeuGlnLysAspSerCysValLeuValTyrr	178
67864	TCATCATGAACCGAATGGCGAGGCAAGAAAGAG..ATTAGCCTTACC	67910
178	roGluMetLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGly	194
67911	GTTGCGCAAAAGTTCTCACCATTTCACCTAGTATTTGAGTGTTCGTATGGG	67960
195	pheGluProGluGlnIleLysThrAspGluGlnGluLeuValGluAlaPh	211
67961ATAAAAGTG...GAGATTGGAGTGTAGAGGTCCT	67992
211	eglGluMetIleLysAsnLeuPheSerLeuProIleAspValProbes	228
67993	TTGAGAGGCGTTTAGAAGGTGTTTCCGCTTGCCGGTTGAGTTCCCTGTT	68042
228	ergLysLeuTyrrArgGlyLeuArgAlaArgAsnPheIleHisSerLysIle	244
68043	CTAAGTTGCTTAGAGCAAGAAGCAAGGCTAGAGATAGAGACATTCTA	68092
245	GluGlnAsnIleArgLysLysIleGlnAspAspAsnGluAsnGluGln	261
68093	GTTGGGAAAGCTCAGGAGAGCGGAAGAGAAATGGAATAAGAGGAGCTGA	68142
261	nLysTyrlLysAspAlaLeu..GlnLeuLeuIleGlnAsnSerArgArgS	277
68143	GAAAGCCAACAACAACGCTTTTCTCGAGGCTAGTAGAA.....	68179
277	erAspGluProPheSerLeuGlnAlaMetLysGluAlaIleThrGlu...	292
68180GAGTTGATCAAAAGGTGTGATTAACCGAAGAA	68209
293LeuLeuPheGlyGlnHisGluThrTh	301
68210	GAGTTGCTAGATATATAGTTTTCCTAGTGTGCTGAGCTCATGACACAAC	68259

[illegible]

seq_documentation_block:			
LOCUS	SLU54770	1395 bp	mrna
DEFINITION	Solanum lycopersicum cytochrome P450 homolog (Dwarf1) mRNA, complete cds.		
ACCESSION	U54770		
VERSION	U54770.1	GI:1421740	

[illegible]

Align seg 1/1 to: AAV09251 from: 1 to: 1850

```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
  |||
88 ATGGGGCTGTACACCCCTATGCTACCTTCTCTGACCATCGCTGTACC 137
17 ovalleuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
  |||
138 CGTTTACTCTTCTCGCCCGCGGTGAGCTGTGGAGATGTTAATGATCC 187
34 rgarValAspProAsnCysArgSerProLeuProGlyThrMetGly 50
  |||
188 GACGAGTCGATCCGAACCTGCAGAGATCTTACCGCAGGTACCATGGGC 237
51 LeuProPheIleGlyGluThrLeuGluLeuIleLeuGluArgArgLysPh 67
  |||
238 TTGCCGTCATTTGAGAAACGCTCCAGCTGATCTCCAGAGAGAGAGATT 287
67 eleuArgMetLysArgGluLysTyrGlyCysIleTyrLysThrHisLeuP 84
  |||
288 TCTGCCGATGAACGGCAGAAATACGGGTGATCTACAGAGCCACCTCT 337
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
  |||
338 TCGGGAACCCGACTGTGAGGGTGTGAGGAGCTGATATGTGAGGCGAGAT 387
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
  |||
388 CTGCTGGGCGAACAAGCTGTGCTGTCTGTCAGTGCCAGCATCAGTGAG 437
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
  |||
438 AACCATCTGGGCTCTGACACCCCTCTCCAATGTCCATGTGAGTTCAACACA 487
134 ysAsnLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
  |||
488 AAAACAAGAAAAAGCCATTATGAGGGCGCTTCTCTGAGATGCTCTGGAG 537
151 HistyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
  |||
538 CACTACATTCCTCGTATCCAGCAGAGGAGGAGGAGCCCATACAGGAATG 587
167 pleuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeu 184
  |||
588 GCTGCAAAAAGACTCTCGCTGCTGTTATCCAGAAATGAAGAAACTCA 637
184 etPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
  |||
638 TGTTCGATAGCTATGAGAAATCTGCTGTTGTTGAACCAAGAGCAATA 687
201 LysThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleLysAs 217
  |||
688 AAGACGACGACGACGACGACGTTGGAAGCTTTTGAAGAAATGATCAAAA 737
217 nleuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyL 234
  |||
738 CTGTTCTCTGCTGCAATCGACGTTCTTCACTGCTGTACAGGGGTT 787
234 euArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArgLys 250
  |||
788 TGAGGCGACGCAATTTCATTCCTCCAAATTTGAGGAAAAACATCAGGAAG 837
251 LysIleGlnAspAspAsnGluAsnGluGlnLysTyrLysAspAlaLe 267
  |||
838 AAAATTCAGATGACGACATGAAAGCAAGCAAGAAATACAAAGAGCGCCT 887
267 uGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerLeuG 284
  |||
888 TCAGCTGTGATCGAAGACAGCAGAAAGAGTGAACCTTTAGTTGTC 937
284 InAlaMetLysGluAlaAlaThrGluLeuLeuPheGlyHisGluThr 300
  |||
938 AGGCGATGAAGAAGACAGCTACAGAGCTTCTATTTGAGAGTCATGAACC 987

```

```

301 ThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGl 317
  |||
988 ACCGCCAGCAGCTGCAACCTCACTTGTGATGTTTCTGGCTGTGAACACAGA 1037
317 uValValGlnLysValArgGluGluValGlnGluLysValGluMetGlyM 334
  |||
1038 AGTGTGCAGAAAGCTCAGAGAGAGGTTTCAGAGAGAGGTTGAAATGGGCA 1087
334 etTyrThrProGlyLysGlyLeuSerMetGluLeuLeuAspGlnLeuLys 350
  |||
1088 TGTATACACCTGGAAGGGCTTGAGTATGAGCTGTTGACCAGCTGAAG 1137
351 TyrThrGlyCysValIleLysGluThrLeuArgIleAsnProProValPr 367
  |||
1138 TACACTGATGTGTGATTAAGAGACTCTTGAATGAATGATGTTACCAAA 1187
367 ogIlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnI 384
  |||
1188 CGAGGATTCAGAGTCGACCTCAAAACCTTGAATGAATGATGTTACCAAA 1237
384 leProLysGlyTTrpAsnValIleTyrSerIleCysAspThrHisAspVal 400
  |||
1238 TTCTTAAGATGGAACGTCAATTACAGCATCTGTGACACGACGAGATG 1287
401 AlaAspValPheProAsnLysGluGluPheGlnProGluArgPheMetSe 417
  |||
1288 GCCGACGCTCTTCCAAACAAAGAGAGAGTTCCAGCCGAGAGATTCATGAG 1337
417 rLysGlyLeuGluAspGlySerArgPheAsnTyrIleProPheGlyGlyG 434
  |||
1338 CAAAGGTCTGAGAGAGCGGCTCCAGGTTTAATCAATCCCTTCGAGAGAG 1387
434 lySerArgMetCysValGlyLysGluPheAlaLysValLeuLeuLysIle 450
  |||
1388 GATCCAGGATGTGTGTGGCAAGAGATTGCCAAAGTGTACTCAAGATC 1437
451 PheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyPr 467
  |||
1438 TTTTGTAGTTGAGTTAACGACGATTTGCAATGTGATCTCTCAAAACGAGC 1487
467 opProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuProT 484
  |||
1488 CCCGACAATGAAACAGCGCCGACTATTATCCAGTGCAGCAATCTCCCTA 1537
484 hrLysPheThrSerTyrValArgAsn 492
  |||
1538 CCAAAATTCAGTATGTCAGAAAT 1563

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12203

seq_documentation_block: ID AAV12203 standard; cDNA; 1850 BP.

```

XX AC AAV12203;
XX AC AAV12203;
XX DT 22-JUN-1998 (first entry)
XX DE Zebrafish retinoid metabolising protein ZP450RAI cDNA.
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; zebrafish; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; ss.

```

XX OS Danio rerio.

```

XX FH key Location/Qualifiers
XX FT CDS 88..1566
XX FT /*tag= a

```

PN WO9749815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI; 1998-077178/07.
DR P-PSDB; AAW12203.
XX
PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 15; Page 52-54; 110pp; English.

This cDNA clone codes for a novel zebrafish retinoid metabolising protein (see AAW44159), designated zp450RAI. The encoded protein is a retinoid oxidase that has the ability to hydroxylate retinoic acid at the 4 position of the beta-ionone ring and is inducible in epithelial cells exposed to retinoic acid. To isolate the clone, transcripts present in fin tissue regenerating in the presence or absence of retinoic acid were compared using a differential display PCR technique, and an isolated clone (see AAV12213) was used to screen a D. rerio 6-18 hr embryo cDNA library. Human (see AAV12204) and mouse (see AAV12205) P450RAI clones have also been identified. The isolated clones can be used in the recombinant production of P450RAI proteins. Antisense nucleic acids can be used in a claimed method for inhibiting retinoic acid hydroxylation for the treatment of cancer, actinic keratosis, oral leukoplakia, secondary tumours of the head and/or neck, non-small cell lung carcinomas, basal cell carcinomas, acute promyelocytic leukaemia, skin cancer, and premalignancy associated with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter sequences (see AAV12206-08) are also claimed.

xx Sequence 1850 BP; 519 A; 391 C; 454 G; 486 T; 0 other;
SQ

alignment_scores:

quality: 2563.00 Length: 492

Ratio: 5.209 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-668-482-2 x AAV12203 ...

Align seg 1/1 to: AAV12203 from: 1 to: 1850

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
 88 ATGGGGCTGTACACCTTATGGTCACCTTCTGTGCACCATCGTGTACC 137
 17 oValLeuLeuPheLeuAlaValIlysLeuTrpGluMetLeuMetIleA 34
 138 CGTTTACTCTTTCTGCGCCGCGGTGAAGTTGTGGAGATGTTATGATCC 187
 34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
 188 GACGAGTCGATCCGAACTCGAGAAGTCCTCTACCGCAGGTACCATGGGC 237
 51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
 238 TTGCCGTTCAATTGGAGAAACGCTCCAGCTGATCCTCCAGAGAAAGATT 287
 67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84

288	TC	TGCGCATGAAACGCGACGAAATACGGGTGCATCTTACAAGACGCACCTCT	337
84	he	gIAsnPrOthrValAlaArgValMetGlyAlaAspAsnValArgGlnIle	100
338	TC	GGGAACCCGACTGTCAGGGTGATGGAGCTGATATGTGAGGCAGATT	387
101	Leu	LeuGlyGlnHisLysLeuValSerValGlnTrpProAlaSerValAr	117
388	CT	GCTGGCGCACACAAGCTGGTGTCTGTTCAGTGGCCAGCATCAGTGAG	437
117	gThr	IleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL	134
438	AA	CCATCCTGGGCTCTGCACACCCCTCTCCAATGTCCATGGAGTTCACACA	487
134	ysAsn	LysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu	150
488	AAA	CAAGAAAGCCATTATGAGGCGCTCTCTCGAGATGCTCTGGAG	537
151	His	TrpIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr	167
538	CA	TACATATCCCGTGATCCAGCAGAGGTGAAGAGCCCATACAGGAATG	587
167	pleu	GlnLysAspSerCysValLeuValTyrProGluMetLysLysLeu	184
588	GCT	CAAAAAGACTCCCGCGTGCTGTTATCCAGAAATGAAGAACTCA	637
184	etp	heArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle	200
638	TGT	TTCGGATAGCTATGACAAATCCCTGCTTGTGTTGAACCAAGCAATA	687
201	Lys	ThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleLysAs	217
688	AAG	ACGACGACGACAGCAAGACTGTGTGAAGCTTTTGAGGAATGATCAAAA	737
217	nleu	PheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyL	234
738	CTT	GTTCTCTTGCCAAATGCAGCTTCTTTCAGTGGTCTGTACAGGGGTT	787
234	eua	ArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArgLys	250
788	TCA	GGGCACGCAATTTCATTCACCTCCAAAATTGAGGAAAAATCAGGAAG	837
251	Lys	IleGlnAspAspAsnGluAsnGluGlnLysTyrLysAspAlaLe	267
838	AAA	ATTCAGATGACGACCAATGAAGAACAGAAATACAAAGACGCCCT	887
267	uGln	LeuLeuIleGluAsnSerArgArgSerAspGluProPheSerLeug	284
888	TC	ACGTGTGATCGAGACAGCAGAGAAGTGACGAACCTTTTACTTGC	937
284	I	nalMetLysGluAlaIleThrGluLeuLeuPheGlyGlyHisGluThr	300
938	AG	CGCATGAAGAAGCAGCTACAGAGCTTCTATTGGAGGTGATGAACC	987
301	Thr	AlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGl	317
988	ACC	CGCAGCCTCAACCTCATTGTTCATGTTTCTGGGCTGACACACA	1037
317	uVal	ValGlnLysValArgGluGluValGlnGluLysValGluMetGlyM	334
1038	AG	TGTCAGAAAGTTCAGAGAGAGGTTCAAGAGGTTGAATGGCA	1087
334	et	TyrThrProGlyLysGlyLeuSerMetCulLeuLeuAspGlnLeuLys	350
1088	TGT	ATACACCTGAAAGGCTTGAGTATGAGACTGTTGGACCAGCTGAAG	1137
351	Tyr	ThrGlyCysValIleLysGluThrLeuArgIleAsnProProValPr	367
1138	TAC	ACTGGATGTGATTAAAGAGACTCTTAGAATCAACCCCTCCTGTCC	1187
367	oGly	GlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnI	384
1188	CG	GAGATTCAAGTCCGACTCAAAACCTTTGAATTGAATGGTTACCAA	1237

```

384 leProlysglyTrpAsnValIleTyrSerIleCysAspThrHisAspVal 400
|||||
1238 TTCCTAAAGATGGAACGTCATTTACAGCATCTGTGACACGACGATGTG 1287
401 AlaAspValPheProAsnLysGluGlnPheGlnProGluArgPheMetSe 417
|||||
1288 GCCGACGCTCTTCCAAACAAGAGAGTTCCAGCCGAGAGATTCATGAG 1337
417 rLysGlyLeuGluAspGlySerArgPheAsnTyrIleProPheGlyGlyG 434
|||||
1338 CAAAGGCTGTGAGGACGGGTCCAGGTTTAACCTACATCCCCCTTCGAGGAG 1387
434 lYSerArgMetCysValGlyLysGluPheAlaLysValLeuLysIle 450
|||||
1388 GATCCAGGATGTGTGGGCAAGAGTTCGCCAAAGTGTACTCAAGATC 1437
451 PheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyPr 467
|||||
1438 TTTTGTAGTTGAGTTAACGACGACATTGCAATTGGATTCTCTCAACGAGACC 1487
467 oProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuProT 484
|||||
1488 CCCGACAATGAATAACAGGCCCGCACTATTACCCAGTGACACAATCTCCCTA 1537
484 hrLysPheThrSerTyrValArgAsn 492
|||||
1538 CCAAAATTCAC TAGTTATGTCAGAAAT 1563
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09247

seq_documentation_block:

ID AAV09247 standard; cDNA; 1494 BP.

AC AAV09247;

DT 07-JUL-1998 (first entry)

DE Human cytochrome P450RAI cDNA sequence.

KW Retinoid regulated gene; cytochrome P450 gene; enzyme;

KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter; ss.

OS Homo sapiens.

Key Location/Qualifiers

FT CDS 1..1494

FT /tag= a

FT /product= "Cytochrome P450RAI"

PN WO9749832-A2.

PD 31-DEC-1997.

PF 23-JUN-1997; 97WO-CA00488.

PR 01-OCT-1996; 96US-0724466.

PR 21-JUN-1996; 96US-0667546.

PA (TOOH) UNIV QUEBENS KINGSTON.

PI Petkovich PW;

DR WPI; 1998-077193/07.

DR P-PSDB; AAW37734.

PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis

PS Example 4; Pages 56-58; 113pp; English.

CC This nucleotide sequence encodes the human cytochrome P450RAI. Its
CC expression is dependent on the presence of retinoic acid (RA). The
CC retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RAI is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen
CC using promoters and nucleotide sequences to identify drugs which are
CC useful for reducing the catabolism of RA.

SQ Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other;

alignment_scores: Quality: 1738.00 Length: 497
 Ratio: 4.080 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.008

alignment_block:
US-09-668-482-2 x AAV09247 ..

Align seg 1/1 to: AAV09247 from: 1 to: 1494

```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||
1 ATGGGCTCCCGCGCGCTGCTGCCAGTGCGCTCTGCACCTTCGCTGCC 50
17 oValLeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
|:::|
51 GCTGCTGCTCTTCTCTGCTGCGATCAAGCTCTGGGACCTGTACTGCGTGA 100
51 rGargValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
|||
101 GCGGCGCGGACCGCAGTGTGCCCCCTCCCATTTGCCCGGAGCTATGGGC 150
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgLysPh 67
:::|
151 TTCCCTTCTTTGGGGAACCTTGACAGATGTACTGACGCGGAAGATT 200
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
|:::|
201 CCTGCAGATGAAGCGCAGAGAAATACGGCTTCATCTACAGACGCACTGT 250
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|:::|
251 TCGGGCGGCGCCACCGTAGGGGTGATGGCGCGGACAAATGTGCGGCATC 300
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|:::|
301 TTGCTCGGAGACGACCGGCTGTGTGCTGCACGTGCCACGCTGCTGCG 350
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|:::|
351 CACCATTTCTGGGATGTGCTGCTCTCTTAACCTGCACGACTCCTCGACA 400
134 yAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
|:::|
401 AGCAGCGCAAGAAGTGATATGCGGCCCTTCAGCCGCGAGGACTCGAA 450
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
|:::|
451 TGCTACGTGCCGCGTGATCACCGAGGAAGTGGCGACGACCTGGAGCAGTG 500
167 pLeuGlnLysAspSerCys.....ValLeuValTyrProGluM 180
|:::|
501 GCTG.....AGTCGCGCGGAGCGCGGCTCTGCTGTACCCCGAGG 541
180 eTylsLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
|:::|
542 TGAAGCGCCTCATGTTCGAATGCGCATGCGCATCTACTGGGCTGCGAA 591
```


197 ProGluGlnIleLysThrAsp.....GluGlnGluLeuValGluAl 210
||| |||::: ||| |||||:::|||||||
592 CCC...CAACTGGCGGGCGGAGCTCCGAGCAGACTGTGGAGGC 638
210 apheGluGluMetIleLysAsnLeuPheSerLeuProIleAspValProp 227
||||||| |||:::||||||| ||||| ||||| ||||| |||||
639 CTTCGAGGAATGACCCGCAATCTTCTCGCTGCGCCCATCGACGTGCCCT 688
227 heSerGlyLeuTyrArgGlyLeuArgAlaArgAsnPhelIleHisSerLys 243
||||||| |||||:::||||||| |||||:::|||||||:::|
689 TCAGCGGGCTGTACCGGGCATGAGCGCGGAACCTCATTCACGCGCGC 738
244 IleGluGluAsnIleArgLysIleGlnAsp..AspAspAsnGluAs 259
|||||||:::||||||| |||||:::|
739 ATCGAGCAGAACATTTCGCCCAAGATCTGCGGGCTGCGGGCATCCGAGGC 788
259 nGluGlnLysTyrLysAspAlaLeuGlnLeuIleGluAsnSerArga 276
||| ||||| ||||| ||||| ||||| |||||:::|
789 GGGCCAGGGCTGCAAGACGCGCTGACAGCTGTGATCGAGCACTCGTGGG 838
276 rgSerAspGluProPheSerLeuGlnAlaMetLysGluAlaIleThrGlu 292
::| |||:::|||||||:::|||||||:::|
839 AGAGGGAGAGCGGCTGACATGACAGCACTAAAGCAATCTTCAACCGAA 888
293 LeuLeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerLeuVa 309
||||||| ||||| ||||| ||||| ||||| |||||:::|
889 CTCTCTTTGGAGGACACGAAACACGCGCCAGTGACGACCATCTCTGAT 938
309 lMetPheLeuGlyLeuAsnThrGluValValGlnLysValArgGluIuv 326
: ||||| |||||:::|||||||:::||||||| |||||
939 CACTTACCTGGGGCTCTACCCACATGTTCTCCAGAAAGTGCAGAAAGAGC 988
326 aGlnGluLysValGluMetGlyMetYrThrProGlyLysGlyLeuSer 342
:::|:::| |||:::|
989 TGAAGAGTAAGGTTTACTTTGCAAGAGCAATCAAGACAACAAGTTGGAC 1038
343 MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGluTh 359
|||||||:::||||||| ||||| ||||| ||||| |||||
1039 ATGGAATTTTGGACAACACTTAATACATCGGGTGTATTAAAGAGAC 1088
359 rLeuArgIleAsnProProValProGlyGlyPheArgValAlaLeuLysT 376
|||||||:::||||||| ||||| ||||| ||||| |||||
1089 CCTTCGACTGAATCCCCCAGTTCACAGAGGCTTTCGGGTGCTCTGAAGA 1138
376 hrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyr 392
||||||| ||||| ||||| ||||| ||||| ||||| |||||
1139 CTTTGAATTAAATGGATACAGATTCCCAAGGCGCTGGAATGTTATCTAC 1188
393 SerIleCysAspThrHisAspValAlaAspValPheProAsnLysGluI 409
||||||| ||||| ||||| ||||| ||||| ||||| |||||
1189 AGTATCTGTGATCATGATGATGGAGAGATCTTCAACCAAGGAAGA 1238
409 upheGlnProGluArgPheMetSerLysGlyLeuGluAspGlySerArgp 426
|||||||:::|||||||:::| |||||:::|
1239 ATTTAATCCTGACCGAATTCAGTCTCTCACCACAGAGATGCATCCAGGT 1288
426 heAsnTyrIleProPheGlyGlyLysSerArgMetCysValGlyLysGlu 442
||:::| ||||| ||||| ||||| ||||| ||||| |||||
1289 TCAGCTTCATTCATTTGAGAGAGCGCTTAGAGCTGTGTAGGCAAAAGAA 1338
443 pheAlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCy 459
|||||||:::||||||| ||||| ||||| ||||| |||||
1339 TTTGCAAAATTTCTTCAAAATATTTACAGTGGAGCTGGCCAGGCATGTG 1388
459 sAsnTrpIleLeuSerAsnGlyProProThrMetLysThrGlyProThri 476
||:::| ||| ||||| ||||| ||||| ||||| |||||:::|
1389 TGACTGGCAGCTTCTAAATGACCTCCTACAAATGAAAAACAGTCCACCG 1438
476 leTyrProValAspAsnLeuProThrLysPheThrSerTyr 489
:::| ||||| ||||| ||||| ||||| ||||| |||||:::|
1439 TGTATCTCTGTGACAATCTCCCTGCAAGATTCAACCCATTTTC 1479
seq_name: /sids1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12204

seq_documentation_block:
ID AAV12204 standard; cDNA; 1494 BP.
XX
AC AAV12204;
XX
DT 22-JUN-1998 (first entry)
XX
DE Human retinoid metabolising protein hp450RAI cDNA.
XX
KW Retinoid metabolising protein; p450RAI; retinoid oxidase;
KW retinoic acid; human; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ss.
XX
OS Homo sapiens.
XX
PN WO97/49815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
XX Beckett BR, Jones G, Petkovich PM, White JA;
PI WPI; 1998-077178/07.
DR P-PSDB; AAW44160.
XX
PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 15; Page 55-57; 110pp; English.
XX
CC This cDNA clone codes for a novel human retinoid metabolising
CC protein (see AAW44160) designated hp450RAI. The encoded protein is a
CC retinoid oxidase that has the ability to hydroxylate retinoic acid
CC at the 4 position of the beta-ionone ring and is inducible in
CC epithelial cells exposed to retinoic acid. To isolate the clone,
CC zebrafish p450RAI was used to search an expressed sequence tag
CC database, and an isolated clone was used to screen a cDNA library
CC generated from an NT2 cell line treated with retinoic acid. The
CC hp450RAI gene has been localised to 10q23-24. Zebrafish, human
CC and mouse p450RAI cDNA clones (see AAV12203-05) are claimed. They
CC can be used in the recombinant production of p450RAI. Antisense
CC nucleic acids can be used in a claimed method for inhibiting
CC retinoic acid hydroxylation for the treatment of cancer, actinic
CC keratosis, oral leukoplakia, secondary tumours of the head and/or
CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis.
CC Promoter sequences (see AAV12206-08) for the p450RAI genes are also
CC claimed.
XX
SQ Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other;

alignment_scores:
Quality: 1738.00 Length: 497
Ratio: 4.080 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.008

alignment_block:
US-09-668-482-2 x AAV12204 ..

Align seg 1/1 to: AAV12204 from: 1 to: 1494


```

1266 TAAATCCGACCGCTTTATAGTCCTCATCCAGAGGATGCTCCCGGTTCA 1315
427 sNTYrIIeProPheGLyGLySerArgMetCysValGlyLysGluPhe 443
1316 GCTTCATTCATTTGGAGGAGGCCCTTCGGAGCTGTGTAGGCAAGAGATT 1365
444 AlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAs 460
1366 GCAAAATTTCTTCTTAAGATATTACAGTGGAGCTGGCTAGGCACTGTGA 1415
460 nTPrlleuSerAsnGlyProProThrMetLysThrGlyProThrIleT 477
1416 TTGGCAGCTTCTAATGACCTCCTACAAATGAAGACAAAGCCCCACTGTGT 1465
477 yrProValAspAsnLeuProThrLysPheThrSerTyr 489
1466 ACCCTGTGACAATCTCCCTGCAAGATTCACCTACTTC 1503
seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12205
```

```

seq_documentation_block:
ID AAV12205 standard; cDNA; 1725 BP.
XX
AC AAV12205;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolising protein mp450RAI cDNA.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 25..1518
FT /tag= a
XX
PN W09749815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI; 1998-077178/07.
DR P-PSDB; AAW44161.
XX
PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 15; Page 62-64; 110pp; English.
XX
CC This cDNA clone codes for a novel mouse retinoid metabolising
CC protein (see AAW44160) designated mp450RAI. The encoded protein is a
CC retinoid oxidase that has the ability to hydroxylate retinoic acid
CC at the 4 position of the beta-ionone ring and is inducible in
CC epithelial cells exposed to retinoic acid. The clone was isolated
CC from a retinoic acid-treated P19 teratocarcinoma cDNA library
CC using human P450RAI cDNA (see AAV12204) as probe. Zebrafish, human
CC and mouse P450RAI cDNA clones (see AAV12203-05) are claimed. They
```

```

CC can be used in the recombinant production of P450RAI proteins.
CC Antisense nucleic acids can be used in a claimed method for
CC inhibiting retinoic acid hydroxylation for the treatment of cancer,
CC actinic keratosis, oral leukoplakia, secondary tumours of the head
CC and/or neck, non-small cell lung carcinomas, basal cell carcinomas,
CC acute promyelocytic leukaemia, skin cancer, and premalignancy
CC associated with actinic keratosis, acne, psoriasis and/or
CC ichthyosis. Promoter sequences (see AAV12206-08) for the P450RAI
CC genes are also claimed.
SQ Sequence 1725 BP; 408 A; 425 C; 475 G; 417 T; 0 other;

alignment_scores:
Quality: 1734.00 Length: 496
Ratio: 4.070 Gaps: 4
Percent Similarity: 85.887 Percent Identity: 67.742

alignment_block:
US-09-668-482-2 x AAV12205 ..

Align seg 1/1 to: AAV12205 from: 1 to: 1725
1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
25 ATGGGGCTCCCGCGCGCTGCTGCCAGTGCCTCTGCACCTCTGCTGCC 74
17 oValLeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
75 GCTGCTGCTCTTCTTGGGGAACATTCAGATGGTGTGCTTACAGCGAAGTT 124
34 rgArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
125 GCAGCGCGATCGCAGCTGCGCCCTCCCTTGCCCCCGGTACCATGGGC 174
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
175 TTCCCATTTCTTGGGGAACATTCAGATGGTGTGCTTACAGCGAAGTT 224
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
225 TCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAGACGCATCTGT 274
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
275 TTGGCGCGCCACACGGTGGGTGATGGCGCGGATATATGTGGCGCATC 324
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
325 TTGCTGGAGAGACACGGGTGTGTGGTGCACCTGGCCCGCTGGGTGCG 374
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
375 CACCATCTGGGCGCTGGCTGCCTCTCCAACTGCACAGATTCTCCGACAC 424
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
425 AGCAGCGAAGAAGGTGATTATGCAGGCTTCAGCCGCGAGGCACTCCAG 474
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
475 TGCTACGTGCTCGTGTATCGCTGAGGAAGTCAGCAATTGTCTGGAGCAGTG 524
167 pleuGlnLysAspSerCys.....ValLeuValTyrProGluM 180
525 GCTA.....AGTGGCGGCGAGCGCGGCTCTGCTTACCCCGAGG 565
180 etLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
566 TGAAGCGCCTCATGTTCGCATCGCCATGCGCATCTGCTGGCTGCGAG 615
197 ProGluGlnIle.....LysThrAspGluGlnGluLeuValGluAlaph 211
||| |||||||||:::|||||||
```



```

84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 TGGGGCGGCCACCGTACCGGTGATGGCGCGGACAAATGTGGCGGCATC 345
101 leuLeuGlyGlnHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
346 TTGCTCGGAGACACCGCGTGTGCTCCACTGCGCAGCGTGGTGG 395
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
396 CACCAATTCTGGGATCTGGCTGCTCTTAACCTGCACGACTCTCCGACA 445
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
446 AGCAGCCCAAGAGGTGATGATGATGCGGCTTCAGCCGCGAGGCACTCGAA 495
151 HisTrIleProValIleGlnGlnValLysSerAlaIleGlnGlnLTr 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
496 TGCTACGTGCGCGGTGATCACCGAAGAGTGGGACAGACCTGGAGACAGTG 545
167 pleuGlnLysAspSerCys.....ValLeuValTyrProGluM 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
546 GCTG.....AGCTGCGCGAGCGCGGCTCTCTGCTACCCCGAGG 586
180 etLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
587 TGAAGCGCTCATGTTCCGAATGCCATGCCATCTACTGGCTGCCGAA 636
197 ProGlnIleLeuLysThrAsp.....GluGlnLeuValGluAl 210
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
637 CCC...CAACTGGCGGCGGAGCGGACCTCCGAGCAGACGCTTGTGGAGGC 683
210 apheGlnLysMetIleLysAsnLeuPheSerLeuProIleAspValProp 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
684 CTTCGAGGA.ATGACCGCGCAATCTCTCTGCTGCCATCGACGTGCCCT 732
227 heSerGlyLeuTyrArgGlyLeuArgAlaArgAsnPheIleHisSerLys 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
733 TCAGCGGCTGTACCGGGCATGAAGCGCGGAACCTCATTCACGCGCGC 782
244 IleGlnLysAsnIleArgLysLysIleGlnAsp..AspAspAsnGluAs 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
783 ATCGAGCAGAACATTCGCGCAAGATCTGCGGCTGCGGCGCATCCGAGGC 832
259 nGlnLysLysTyrLysAspAlaLeuGlnLeuIleGlnLysSerArgA 276
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
833 GGGCAGGCGCTGCAAAAGACGCGCTGCAAGCTGTGATCGAGCAGCTCGTGG 882
276 rgSerAspGluProPheSerLeuGlnAlaMetLysGlnAlaAlaThrGlu 292
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
883 AGAGGGGAGAGCGGCTGACATGACGACACTAAAGCAATCTCAACCGAA 932
293 leuLeuPheGlyGlyHisGlnThrThrAlaSerThrAlaThrSerLeuVa 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
933 CTCCTCTTTGGAGGACAGCAAGACGCGCAGTGCAGCCACATCTCTGAT 982
309 lMetPheLeuGlyLeuAsnThrGluValValGlnLysValArgGlnLuv 326
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
983 CACTTACCTGGGCTCTACCCACATGTCTCCAGAAAGTGGCAGAGAAGGC 1032
326 aIGlnLysValGlnMetGlyMetTyrThrProGlyLysGlyLeuSer 342
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1033 TGAAGAGTAAGGTTTACTTTGCAAGACAAATCAAGACAAAGATTGGAC 1082
343 MetGlnLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGlnTh 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1083 ATGGAATTTTGAACAACACTTAATACATCGGGTGTATTATAAGAGAGAC 1132
359 rLeuArgIleAsnProProValProGlyGlyPheArgValAlaLeuLysT 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1133 CCTTCGACTGAATCCCCCAGTTCCAGAGGGGTTTCGGGTTGCTCTGAAGA 1182

```

```

376 hrPheGlnLeuAsnGlyTyrGlnIleProLysGlyTyrPAsnValIleTyr 392
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1183 CTTTGAATTAATGATACAGATTCACCAAGGGCTGGAATGTTATCTAC 1232
393 SerIleCysAspThrHisAspValAlaAspValPheProAsnLysGlnL 409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1233 AGTATCTGTGATCATGATGTGGCAGAGATCTTCACCAACAAGAGA 1282
409 upheGlnProGluArgPheMetSerLysGlyLeuGluAspGlySerArgP 426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1283 ATTTAATCCCTGACCGATTCATGCTGCCCTCACCCAGAGATGCATCCAGGT 1332
426 heAsnTrpIleProPheGlyGlyLysArgMetCysValGlyLysGlu 442
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1333 TCAGCTTCATTCATTTGGAGAGAGCCCTTAAGAGCTGTGTAGGCAAGAA 1382
443 pheAlaLysValLeuLeuLysIlePheLeuValGlnLeuThrGlnHisCy 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1383 TTTGCAAAATTTCTTCTCAAAATATTACAGTGGAGCTGGCCAGGCAATTG 1432
459 sAsnTrpIleLeuSerAsnGlyProProThrMetLysThrGlyProThri 476
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1433 TGACTGGCAGCTTTCTAAATGACCTCTTACAATGAAAAACCAAGTCCACCG 1482
476 leTyrProValAspAsnLeuProThrLysPheThrSerTyr 489
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1483 TGTATCTGTGGAACAATCTCCCTGCAAGATTCACCCATTTTC 1523
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12216
seq_documentation_block:
ID AAV12216 standard; DNA; 4164 BP.
XX
AC AAV12216;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolizing protein mp450RAI genomic DNA.
XX
KW Retinoid metabolizing protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
XX
OS Mus musculus.
XX
PN W09749815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI, 1998-077178/07.
XX
PT Retinoid metabolizing protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Disclosure; Page 69-70; 110pp; English.
XX
CC This sequence comprises a genomic clone of a novel mouse retinoid
CC metabolising protein designated mp450RAI. A cDNA clone (see
CC AAV12205) that includes a coding sequence for the full-length
CC mp450RAI protein (see AAV44161) is also provided. The encoded

```

CC protein is a retinoid oxidase that has the ability to hydroxylate
CC retinoic acid at the 4 position of the beta-ionone ring and is
CC inducible in epithelial cells exposed to retinoic acid. Zebrafish,
CC human and mouse P450RAI cDNA clones (see AAV12203-05) are claimed.
CC They can be used in the recombinant production of P450RAI.
CC Antisense sequences can be used in a claimed method for inhibiting
CC retinoic acid hydroxylation for the treatment of cancer, actinic
CC keratosis, oral leukoplakia, secondary tumours of the head and/or
CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter
CC sequences (see AAV12206-08) for the P450RAI genes are also claimed.
XX
SQ Sequence 4164 BP; 997 A; 975 C; 1071 G; 1121 T; 0 other;

alignment_scores:
Quality: 1169.00 Length: 1055
Ratio: 3.005 Gaps: 16
Percent Similarity: 36.872 Percent Identity: 29.858

alignment_block:
US-09-668-482-2 x AAV12216 ..

Align seg 1/1 to: AAV12216 from: 1 to: 4164

```
34 ArgArgValAspProAsnCysArgSerProLeuProPro..... 46
||| ||||| ::: ||||| |||||
25 AGAGGGGTAGACGACGACAGATGCTCCCGCGCCCTCATGCGGGTTG 74
47 .GlyThrMetGlyLeuProPheIleGlyGluThrLeuGlnLeuIleuG 63
||| |||
75 CGGTCTCTCTCTCCACCT.....CCCTCTC 100
63 InArgArgLysPheLeuArgMetLysArgGlnLysTyrGlyCysIleTyr 79
||||| ||||| ||||| ||||| |||||
101 AGCGAGAGGAAGTTCTTCAGATGAAGCGCAGGAAATACGGCTTCATCTAC 150
80 LysThrHisLeuPheGlyAsnProThrValArgValMetGlyAlaAspAs 96
||||| ||||| ||||| ||||| |||||
151 AAGACGCATCTGTTGGGGGGCCACGCGTGGGTGATGGCGCGGATAA 200
96 nValArgGlnIleLeuLeuGlyGlnHisLysLeuValSerValGlnTrp 113
||||| ||||| ||||| ||||| |||||
201 TGTGCGGGCGCATCTGCTGGGAGAGACACCGGTGTGCTCGTGCACTGGC 250
113 roAlaSerValArgThrIleLeuGlySerAspThrLeuSerAsnValHis 129
||||| ||||| ||||| ||||| |||||
251 CCGCGTCGGTGGCAGCACCATCTGGGCGCTGGCTGCTCTCCAACCTGCAC 300
130 GlyValGlnHisLysAsnLysLysLys..... 138
::: ||||| |||||
301 GATTCTCTGCACAGCAGGAAAGAGGTAGGGGTGAGCTGGCACTCCT 350
138 ..... 138
351 TGGCTGGCAGGAGACCTCATCTATGTGTGGGACACGACCTCTCTATCT 400
138 ..... 138
401 GCGGGCGAGGGCTAGTCCATATGTGTGGGACACGACCTCTCTATCT 450
138 ..... 138
451 GAGATCCACTTTAGCTTTCTGCTAGCACGTGGTTAGTCTGGGGGGA 500
138 ..... 138
501 CTGAATTCCTTGAAGGGTACTCGGAAAGCGGAAGGGGGGGGCTGAGG 550
139 .....Ala.II 140
::: |||
551 GAAAGTAGAGATTTAACACTCTCTGCTCTGGGGGGTGTCTCAGGTGAT 600
```

```
140 enetArgAlaPheSerArgAspAlaLeuGlnHisTyrIleProValIleG 157
||||| ||||| ||||| ||||| |||||
601 TATGACAGGCCTTCAGCCGAGGACACTCCAGTGTACTGCCCCGTGATCG 650
157 InGlnGluValLysSerAlaIleGlnGluTrpLeuGlnLysAspSerCys 173
::: ||||| ||||| ||||| ||||| |||||
651 CTGAGGAAGTCAGACAGTTGCTGTGAGCAGTGCGTA.....AGCTGC 691
174 .....ValLeuValTyrProGluMetLysLysLeuMetPheAr 186
::: ||||| ||||| ||||| ||||| |||||
692 GCGGACCGCGGCTCTCGTGTCTACCCCGAGGTGAAGCGCTCATGTTCCG 741
186 gIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle.....L 201
||||| ||||| ||||| ||||| |||||
742 CATGCCATGCGCATCTGCTGGCTGCGAGCCGGTCCAGCGGCGGCG 791
201 ystHisAspGluGlnGluLeuValGluAlaPheGluGluMetIleLysAsn 217
||||| ||||| ||||| ||||| |||||
792 GGGAGGACGACGACGACGCTCGTGGAGGCTTTCGAGGAGATGACCCGCAAT 841
218 LeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg..... 232
||||| ||||| ||||| ||||| |||||
842 CTCTTCTCTCTTCCATTGACGTGCCCTTTAGCGGCGCTGTACCGGGTAAG 891
232 ..... 232
892 GCGGTTTGGGAGTGGAGTAGGGGAACGCAAGCTCGGCGCATCCGCTCA 941
233 .....Gly.LeuArgAlaArgAsnPheIleHi 241
942 CCGCACGCTCTCTCCGCGTCAAGCGGTGAAGCGCGGCACTTATACA 991
241 sSerLysIleGlnGluAsnIleArgLysLysIleGlnAsp..AspAspA 257
::: ||||| ||||| ||||| ||||| |||||
992 CGCGCGCATGCGAGAGAAACATTCGCGCCAAAGATCCGCGGCTTCAGGCTA 1041
257 snGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuIleGluAsn 273
::: ||| ||||| ||||| ||||| |||||
1042 CAGAGCCGGAATGGGGGTGCAAGGACGCGCTGCAGCTCTGTGATGAGCAC 1091
274 SerArgArgSerAspGluProPheSerLeuGln..... 284
||| ::| ||| ||||| ||||| |||||
1092 TCGTGGAGAGGGGAGAGAGGCTGATATGCAGGTGAGAAGCAATTCAA 1141
284 ..... 284
1142 AAGTGCCAGGGCGGGGAGTGCCCTGTGACTTCCAGACACACTTCTG 1191
284 ..... 284
1192 GGGTCTCCAAAGCCCTGTCAAGGCCCGACCTACTTCCAAGTGGCGCGCA 1241
284 ..... 284
1242 TGCTAGGTCTAGAGCTTTCAACCTGTGGGTGTGACCCCTTCACGAGC 1291
284 ..... 284
1292 CAACAACCCCTTTCAGAAGGCTGCGCTAAGAGCATCTGCATATCCGATAT 1341
284 ..... 284
1342 TTACATCAAGAAACATAACAGTAGCAAAATTACCGTTATGAGTAGCAAC 1391
284 ..... 284
1392 AAAGATAATTTATCGTTGGGGGTACACCAACAAGAGAACCGTATTTAA 1441
284 ..... 284
1442 AGGGTGGCATTGGTCTAGAGAGCTGTGGAAGGGGGTGGCTGAGCAATGGG 1491
```

284 284
1492 GAAGATCCCAAGTTCAAAGGGCAAGGCTCATCTACAAAGTTAAAGCGG 1541
284 284
1542 AAGACAGAGATTAAAGGAGTTTGGCTTTTGTGTGCTTTGACTTT 1591
284 284
1592 CTATGAACAACGGAATTTTACCCTTGAAGTCTTCCGTCAATATTTCA 1641
284 284
1642 GGTCAGGTCCTTGTAAACAGTGCATATACTGCACTCAGATCTGTATAAC 1691
285 285
1692 TTCCGTTTATTCCTTAGGCACTAAACAATCGTCAACAGAGCTCCTT 1741
295 eGlyGlyHisGluThrThrAlaSerThrAlaThrSerLeuValMetPheL 312
1742 TGGTGTGTCATGAACCTACAGCGCAGTGTGCGACGTCACCTGATCACTTACC 1791
312 euGlyLeuAsnThrGluValGlnLysValArgGluGluValGlnGlu 328
1792 TAGGACTCTACCCACATGTCCTCCAGAAAGTTGAGAAAGATTAAGAGC 1841
329 LysVal..... 330
1842 AAGGTAGATGATTCTAGAGTTCCCATTTGCCCTAGACATTCCTCTAT 1891
330 330
1892 TAAACACACACACACCCCACTGTATATAAGTTTGCCTGATACACCAG 1941
330 330
1942 TACTATGACAGTGAAGATCTGAGAGCTAGGTGGAGCTGTGGGGAGAGAC 1991
330 330
1992 TCCACCTCGTGAATTTAAAAAGCAGTTGTTGTACTGGCTCTCTCTTG 2041
330 330
2042 GGCAGAATTTGACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 2091
330 330
2092 TCCACACACACACACCACTTTTATAGAGCAAGTTCTCTCTTCC 2141
330 330
2142 CTGACCAAGAACATGATTAATGTGATTAGAGCCAATAGCTGATCAGGGTC 2191
331Glu..MetGlyMetTyrThrProGlyLys.. 339
2192 GCAGTGTGGTGAGGGCTCAGGGTATGACCCTTTATATACCTGATTAAGCA 2241
339 339
2242 ACATTTGTCTGATAATGGGTTTAGGCTGAGGAAGTGGGAAAGAGGCC 2291
339 339
2292 ATCAGGCCATCAGTCTTCCCTTTTATCCTCTCCCATCCAGAGCCTTC 2341
339 339
2342 AGGTTTAGTTAACAGGTGAGTCCCTGCTGGGCTGACTTTTTTTTGGAGTG 2391
339 339

2392 CCCAGGATCCATCACTCACTTTTATCTGTTCATAGGCTTACTTT 2441
340GlyLeuSerMetGluLeuLeuAspGlnLeu 349
2442 GCAAGACAATCAAGACAACAAGTTAGACATGGAACCTTTGGCACAGCTT 2491
350 LysTyrThrGlyCysValIleLysGluThrLeuArgIleAsnProProVa 366
2492 AAATACACTGGGTGTGTCATTAAAGAGACCCTGCGATTGAATCCTCCGGT 2541
366 lProGlyGlyPheArgValAlaLeuLysThrPheGlu..... 378
2542 TCCAGAGGGTTTCGGGTGCTCTGAAGACTTTTGAAGCTGAATGTAGTG 2591
378 378
2592 CACCTCCTGTCCCAACCCCAAGCCCTGTCACGTCACCTGTGATGC 2641
378 378
2642 TGTGAGCATCAGCTGCCAGAGCAGTGGCTCACTGCCCTTGACAGTGC 2691
378 378
2692 CTGCTCCTATGTGTAAGTGGGAACCAATTTGCTCTCTCTTAATGCCAT 2741
378 378
2742 CCATGCTAGTAATGACTTTTGTGTGCAAGCTCAGGGCCGGGATTGTC 2791
379LeuAsn..GlyTyrGlnIleProLys 386
2792 AATTCTTAGGATTTTCTTTTAAACAGGATACCAGATCCCAAG 2841
387 GlyTyrAsnValIleTyrSerIleCysAspThrHisAspValAlaAspVa 403
2842 GGCTGGAATGTATTATACAGTATCTGTGACACACCAAGATGTGGCAGATAT 2891
403 lPheProAsnLysGluGluPheGlnProGluArgPheMetSerLysGlyL 420
2892 CTTCACTAACAAAGAGGAATTTATCCCGACCGCTTATAGTGCCTCATC 2941
420 euGluAspGlySerArgPheAsnTyrIleProPheGlyGlySerArg 436
2942 CAGAGATGCTTCCCGGTTCACTTCATTCATTGAGAGAGGCTTCGG 2991
437 MetCysValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeuVa 453
2992 AGCTGTAGGCCAAAGAGTTGCCAAAATTTCTTAAAGATATTACAGT 3041
453 lGluLeuThrGlnHisCysAsnTyrIleLeuSerAsnGlyProProThrm 470
3042 GGAGCTGGCTAGGCACCTGTGATTGGCAGCTTCTAATGAGCCTCTACAA 3091
470 etLysThrGlyProThrIleTyrProValAspAsnLeuProThrLysPhe 486
3092 TGAAGACAAGCCCACTGTGTACCTGTGACAAATCTCCCTGCAAGATT 3141
487 ThrSerTyr 489
3142 ACCCACTTC 3150
seq_name: /SIDSl/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12214
seq_documentation_block:
ID AAV12214 standard; DNA; 2677 BP.
XX
AC AAV12214;
XX
XX 22-JUN-1998 (first entry)
DT
XX Human retinoid metabolising protein hp450RAI genomic DNA.
DE


```

XX Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; human; inhibitor; antisense; cancer;
KM actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KM acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
XX
OS Homo sapiens.
PN WO9749815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PA 21-JUN-1996; 96US-0667546.
XX
( TOOH ) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
DR WPI; 1998-077178/07.
XX
PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
PS Disclosure; Page 68; 110pp; English.
XX
This sequence comprises a genomic clone of a novel human retinoid
CC metabolising protein designated hp450RAI. Another genomic clone
CC (see AAV12215) is also provided, as well as a cDNA clone (see AAV12204)
CC that includes a coding sequence for the full-length hp450RAI
CC protein (see AAW44160). The encoded protein is a retinoid oxidase
CC that has the ability to hydroxylate retinoic acid at the 4 position
CC of the beta-ionone ring and is inducible in epithelial cells
CC exposed to retinoic acid. The hp450RAI gene has been localised to
CC 10q23-24. Zebrafish, human and mouse P450RAI cDNA clones (see
CC AAV12203-05) are claimed. They can be used in the recombinant
CC production of P450RAI. Antisense nucleic acids can be used in a
CC claimed method for inhibiting retinoic acid hydroxylation for the
CC treatment of cancer, actinic keratosis, oral leukoplakia, secondary
CC tumours of the head and/or neck, non-small cell lung carcinomas,
CC basal cell carcinomas, acute promyelocytic leukaemia, skin cancer,
CC and premalignancy associated with actinic keratosis, acne,
CC psoriasis and/or ichthyosis. Promoter sequences (see AAV12206-08)
CC for the P450RAI genes are also claimed.
XX
SQ Sequence 2677 BP; 489 A; 775 C; 868 G; 511 T; 34 other;

alignment_scores:
Quality: 792.00 Length: 673
Ratio: 2.891 Gaps: 13
Percent Similarity: 40.713 Percent Identity: 31.947

alignment_block:
US-09-668-482-2 x AAV12214 ..

Align seg 1/1 to: AAV12214 from: 1 to: 2677

1 MetGlyLeuTyrThrIleMetValThrPheLeuCysThrIleValLeuPr 17
||||| ::||::: |||||||:::|||||
676 ATGGGGCTCCCGGCGGTGTCGCCAGTGCGCTGCACCTTCGTGCTGCC 725
17 oValLeuLeuPheLeuAlaAlaValLysLeuTrpGlutMetLeuMetIlea 34
|:::|||||:::|||||:::|||||:::|:::
726 GCTGCTGCTCTTCCTGCTGCGATCAAGCTGTGGACCTGTACTGCGTGA 775
34 rGaRgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
||| :::||| |||||| |||||

```

```

776 GCGGCCGCGACCGCAGTTGTGTGCCCTCCCATTTGCCCCCCGGGAGCTATISGS 825
51 LeuProPheIleGlyGluThrLeuGlnIleuIleLeuGlnArgArg.LysP 67
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
826 TTCCCTCTCTTTGGGGAAACCTTGCAGATGNTACTNCAGTTAAGGAGGG 875
67 heLeuArg..... 69
|||
876 TGGGGCGGAGACAGCTGCTTCCCGGAGCCCGCGGCTGTGGGCTTCT 925
69 ..... 69
926 GCTGAAGTCGGGGTAGCGCCCCCGGAGGCATGCTATTGCGGCTAGGAG 975
69 ..... 69
976 CAGGCGTGGCGGAGCGCGCGGCTCCCGGMYMCSTCAWGCSCRCMWK 1025
70 .....MetLysArg 72
|||||
1026 TMMCCCTCCGCTYMCSTCCAMAGCGGARSARNKCYKMRGATGAAGCGC 1075
73 GlnLysTyrGlyCysIleTyrLysThrHisLeuPheGlyAsnProThrVa 89
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1076 AGGAATATACGGCTTCACTACAAAGACGATCTGTTCGGCGGCCACCGT 1125
89 largValMetGlyAlaAspAsnValArgGlnIleLeuLeuGlyGluHisL 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1126 ACGGCTGATGGCGCGGACAAATGTGCGGCGCATCTTCTCGGAGAGCACC 1175
106 ysLeuValSerValGlnTrpProAlaSerValArgThrIleLeuGlySer 122
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1176 GCGTGGTGTGCTCCACTGGCCACGCGTCGGTGGCCACCAATCTGGGATCT 1225
123 AspThrLeuSerAsnValHisGlyValGlnHisLysAsnLysLysLys.. 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1226 GCGTGCCTCTTAACCTGCACGACTCTCTCGCACAAAGCAGCCAGAAGGT 1275
138 ..... 138
1276 GGGGCGCAGGAGCGCAGCGCTGAGACAGGAGGGGAGCCCATTTATGACGG 1325
138 ..... 138
1326 GAATTCCGGCTGATGATGCTAGCGCGGGCTAGCAGCTTGAGGTGGCT 1375
138 ..... 138
1376 AGGACCCCTTGCACAGCTCCAGTTAGCTTTCCACAGCTCGAGAGTGCCAT 1425
138 ..... 138
1426 GTGTCTGGCAGGACCTGGGGGTGTCTGGAAGGGGAGCGCGGTAGACGAGAG 1475
139 .....AlaIle 140
:::|||||
1476 GGGCGGATGAGGCTTTTAAACGCTGTCCCTCTCGGAGCTCAGGTGATT 1525
141 MetArgAlaPheSerArgAspAlaLeuGlnHisTyrIleProValIleGln 157
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1526 ATGCGGCGCTTCAAGCCGCGAGGCACTCGAATGCTACGTGCGGTGATCAC 1575
157 ngInGluValLysSerAlaIleGlnIleuTrpLeuGlnLysAspSerCys. 173
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1576 CGAGGAAGTGGCGCAGCAGCCTGAGAGCACTGGCTG.....AGCTGCG 1616
174 .....ValLeuValTyrProGluMetLysLysLeuMetPheArg 186
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1617 GCGAGCGCGGCTCTCTGTGTACCCCGAGGTGAAGCGCTCATGTTCGGA 1666
187 IleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIleLysThrAs 203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1667 ATCGCCATGCGCATCTCTACTGGGCTGCGAACC...CAACTGGCGGCGGA 1713

```


PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147363.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0148171.
PR	10-AUG-1999;	99US-0148319.
PR	11-AUG-1999;	99US-0148341.
PR	12-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	13-AUG-1999;	99US-0149368.
PR	16-AUG-1999;	99US-0149175.
PR	17-AUG-1999;	99US-0149426.
PR	18-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	20-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	23-AUG-1999;	99US-0150566.
PR	25-AUG-1999;	99US-0150884.
PR	26-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160778.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0161040.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.

```

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
      quality: 552.50      length: 485
      ratio: 1.811      gaps: 14
Percent similarity: 62.887      Percent identity: 29.897

alignment_block:
US-09-668-482-2 x AAC47986 ..

Align seg 1/1 to: AAC47986 from: 1 to: 1649

19 LeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleArgAr 35
   |||||:::||||:::||||:::||||:::||||:::||||:::||||:::
79 CTCTCTCTCTCTCTCTCTCCATCGCCGCCGCTTCCTCTCTCTCTCTCCGCCG 128
35 gValAspProAsnCysArgSerProLeuProProGlyThrMetGlyLeuP 52
   ||:::  ::::  |||  ||||| ||||| ||||| ||||| ||||| |||||
129 TACACGTTACCGT..CGAAGGGCTCTCTCCGGGAAGCCTTGGTCTCC 175
52 ropheIleGlyGluThrLeuGlnIleuIle.....LeuGln 63
   ||:::||||||| |||||:::|||||||
176 CTCTGATAGAGAGACTTTTCAGCTGATCGAGACTTACAAACAGAGAAC 225
64 ArgArgLysPheLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLy 80
   ::::  ||:::  ::::  |||||  :::  |||||  :::::
226 CCTGAGCCTTTCATCGACGAGAGATGCCCGGTACGGTTGCTTTCAT 275
80 sThrHisLeuPheGlyAsnProThrValArgValMetGlyAlaAspAsnV 97
   ||||| ||||| ||||| ::: ||||| :::
276 GACGCACTTTTGGTGAACCGACGATTTTCTCAGCTGACCCGGAACGA 325
97 aLArgGlnIleLeuLeuGlyGluHisLysLeuValSerValGlnTrpPro 113
   |||  ::|||  ::|||  |||||  ::  ::|||  :::  ||||
326 ACCGGTTGTCTTTCAGAACGAAAGGAGCTTTTGAGTGTCTTATCT 375
114 AlaSerValArgThrIleLeuGlySerAspThrLeuSerAsnValHisGl 130
   |||||:::  ::::: |||||:::  ::|||  :::  |||
376 GCTTCCATTTGTAAACCTTTTGGGAAACACTCTCTGCTTCTTATGAAGG 425
130 yValGlnHisLysAsnLysLysLysAlaIleMetArgAlaPheSerArgA 147
   |  |||||  :::  :::::  ::|||  ::|||  :::::
426 TTCTTTGCATAAA...CGTATGCACCTCTCAACCATGAGCTTTGCTAAT 472
147 sPalaleu.....GluHisTyrIleProValIleGlnGlnGluValLys 161
   :::::  ::|||  ::::  |||||:::  ||||:  ||||:
473 CTTCATATCATTAAGACCATCTCATGCTTGATATTAACCGGTTAGTCCGG 522
162 SerAlaIleGlnGluTrpLeuGlnLysAspSerCysValLeuValTyrPr 178
   :::::  |||  ::|||  |||||:::
523 TTTAATCTTGATTCCTGG.....TCCTCTCGTGTCTTCCTCATGGA 563
178 ogLueMetLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyP 195
   |||  |||||:::  ||:::  :::::  ::|||  ::|||  ::|||
564 AGAAGCCAAAAAGATTAACGTTTGAGCTAAACGGTGGAAGCAGTTGATGAGCT 613
195 hegluPro.....GluGlnIleLysThrAspGluGlnGluLeu 207
   ||:::  |||  ||:::  :::::  ::|||  ::|||  ::|||
614 TTGATCCAGGGAATGAGATGAGAGTTTAAGGAAGAGATCTTCTTGTGC 663
208 ValGluAlaPheGluGluMetIleLysAsnLeuPheSerLeuProIleAs 224
   ::|||:::  |||  ||||| ||||| ||||| ||||| ||||| |||||
664 ATCGAAGGCTTC.....TTCTCTCTCTCTCTCC 692
224 pValProPheSerGlyLeuTyrArg..GlyLeuArgAlaArgAsnPheI 240
   :::  |||||  |||||  :::::  ::|||  ::|||  ::|||
693 TCTC..TTCTCCACCACTTACCGCAAGCCATCCAAAGCGGGAAGG 739
240 leHisSerLysIleGluGluAsnIleArgLysLysIleGlnAspAspAsp 256
   :::  :::  ||||:  :::::  :::::

```

740 TGGCGGAGCGCTGACGGTGTGTGATGAAAAAG...AGGAGAGAGAG 786
257 AsnGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuIleGluAs 273
:::||||::: ::::: ||||| ||| |||:::
787 GAAGAAGAGCGAGAGAAAAAGAAAGATATGCTTGCGCGTTCCT... 831
273 nSerArgSerAspGluProPheSerLeuGlnAlaMetLysGluAla 290
832GCGGCGATGATGATGATTTCCGATGAAGAGATTGTGACTTCT 874
290 laThrGluLeuLeuPheGlyGlyHisGluThrThraLaserThrAlaThr 306
::: ||||| :::::|||||||::: |||
875 TGGTGGCTTACTTGTGCGCGGTATGAACAACCTCCACGATCATGACT 924
307 SerLeuValMetPheLeuGlyLeuAsnThrGluValGlnLysValAr 323
925 CTGCGCGTCAAAATTCTCACCAGACTCCTTTAGCTCTTGCTCAACTCAA 974
323 gGluGluValGlnGluLysValGluMetGlyMetTyrThrProGlyLysG 340
:||||| :::|||||::: ::||| :::
975 GGAAGAG...CATGAAAAGATTAG...GCAATGAAGAGTGAATCGTATA 1018
340 lLeuSerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIle 356
::|||::: ::: ::::: ::||| |||||:::
1019 GTCTTGAATGAGTGAATTACAAAGTCAATGCCATTACACACATGTGTGTT 1068
357 LysGluThrLeuArgIleAsnProProValProGlyGlyPheArgValAl 373
:::|||||||::: ::: ||| ||||| ||
1069 AATGACGCGCTACGAGTGGCTAACATCATCGCGGTGTTTCAGACGTGC 1118
373 aleuLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnV 390
|::: ||:::|||||:::|||||||:::|
1119 AATGACGCGATGTGAGATCAAGGTTATAAATTCCAAGGCGTGAAG 1168
390 aIleTyrSerIleCysAspThrHisAspValAlaAspValPheProAsn 406
|::: ||::: ::||| ::::: |||
1169 TATTCATCGTTAGACGGTTCATTTAGACCCCAACCACTTCAAGAT 1218
407 LysGluGluPheGlnProGluArgPheMetSerLysGlyLeuGluAspGl 423
::: ||:::|||| ||:::||||:::|
1219 GCTCGCACTTCAACCCTTGGAGATGGCAGAGCAACTCGTTAACGACAGG 1268
423 ySerArgPheAsnTyrIleProPheGlyGlySerArgMetCysValG 440
- ||::: ::: ||||| ||||| |||
1269 CCCTCTAATGTGTACACCGTTTGGTGAGGGGCCAAGCCTATGTCCCG 1318
440 lYlySGluPheAlaLysValLeuLeuLysIlePheLeuValGluLeuThr 456
|| ||:::||||| ||:::||||:::|
1319 GTTACGAGCTGGCTAGGTTGCACACTCTGTGTTCTTCACCGCCTAGTG 1368
457 GlnHisCysAsnTyrPheLeuSerAsnGlyProProThrMetLysThrGl 473
:::||||::: ||::: :::
1369 ACAGGCTTCAGTTGGGTT.....CCTGCAGAGCAAGACAA 1403
473 yProThrIleTyrProValAspAsnLeuProThrLysPheThrSerTyrV 490
:::||||:::| ::::: ||::: ||
1404 GCTGTTTCTTCCAACTACAGAAGCAGAAACGGTACCCGATCTTCG 1453
490 aLArg 491
|::: ||
1454 TGAAG 1458

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAc48714

seq_documentation_block:

ID AAC48714 standard; DNA; 1398 BP.

XX AAC48714;

AC

XX 18-OCT-2000 (first entry)

DT

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 58492.

DE

XX

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0143297.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	28-JUL-1999;	99US-0145919.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	06-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	16-AUG-1999;	99US-0148684.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	30-AUG-1999;	99US-0151080.
PR	31-AUG-1999;	99US-0151438.

PR	01-SEP-1999;	99US-0151930;
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159295.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160815.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161356.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

alignment_scores:		
Quality:	481.50	Length: 484
Ratio:	1.649	Gaps: 13
Percent Similarity:	60.331	Percent Identity: 26.860

alignment_block:
US-09-668-482-2 x AAC48714 . .

Align seg 1/1 to: AAC48714 from: 1 to: 1398

```

1 MetGlyLeuTyrThrLeuMetValI ThrPheLeuCysThrIleValLeuPr 17
   ||||| | : : : : | : : : : | ||||| : : |||
22 ATGGGTCTTCCTTGATCATCGTGCT...CTCTGTCCGCCTCTCCTT.. 66
   ||||| | : : : : | : : : : | ||||| : : |||

17 oValLeuLeuPheLeuAlaIalavalLysLeuTrpGlumetLeuMetIlea 34
    ..... CGATGAATCAGATCGGATAT. 87
    ||| : : : : :

34 rgArgValaspProasnCysArgSerProLeuProProGlyThrmetgly 50
   : : : : | ||||| ||||| ||||| |||||
88 .....ACCAAGAATGCTCTTCCTCCTCGAACCATGGGC 120
   : : : : | ||||| ||||| ||||| |||||

51 leuProPheileGlygluThrLeuglnleuileleuglnargarylspH 67
   ||| : : : : | ||||| | : : : : | ||| : : ||
221 TGGCCAACTCTTTGGCCAACAACCCAGCGATTCTCAACACAGA GCCCAA CTT 170

```

67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
171 CATGAGAAACCAAGACTCCGATACGGAGGTTTCTCAAAATCTCATCTTC 220
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
221 TAGGTTGTCCAAAGTTAAATCTCAATGACTCAGAAAGTAAACAGATACATT 270
101 LeuLeuGlyGlnHisLysLeuValSerValGlnTrpProAlaSerValAr 117
271 TTAAAGAACGAATCAAAAGTTGGTTCTCGTTACCCACAATCGATGCT 320
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
321 TGAATATACTTGGGACTTGTAAACATGGCTGCGGTTACGCGTTCAGCCACC 370
134 ysAsnLysLysLysAlaIleMetLarGAlaPheSerArgAspAlaLeu... 149
371 GGCTTATGAGAGGCTCGCTCTCTCTCATAGCTCGACCATGATGAGA 420
150 GluHisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnG 166
421 GATCATATCTTGCCCTAAAGTTGATCATTCTCATGAGAAGCTATCTTGATCA 470
166 uTrpLeuGlnLysAspSerCysValLeuValTyrProGluMetLysLysL 183
471 GTGGAATGAGCTTGAG...GTTATTGATATCCAAGATAGACCAAAACATA 517
183 euMetPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGln 199
518 TGGCATTTTATCTTCACTGACACAAATCGCTGG...AACTTA 558
200 IleLysThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleL 216
559 AGAAAAACCATTTGTTGAGAATTCAAAACCTGCATTCTTCAAGCTTGTGT 608
216 sAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgG 233
609 TGGGACTTTATCCGTTCCGATGATCTTCCGGGCAAAATTATCGTTGGC 658
233 lLeuArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArg 249
659 GAATCCAAGCAAGAAATAACATT.....GATAGGTGCTA 693
250 LysLysIleGlnAspAspAsnGluAsnGluGlnLysTyrLysAspAl 266
694 AGAGAGCTGATGCAGAACGTTAGAGATTCTGGAGAAACAATTACACAGACAT 743
266 aLeuGlnLeuLeuIle....GluAsnSerArgArgSerAspGluProP 281
744 GTTAGGTTACTTGATGAAGAGGAGGTAACCGA.....TACCCGT 784
281 heSerLeuGlnAlaMetLysGluAlaAlaThrGlnLeuLeuPheGlyGly 297
785 TAACCGATGAGAGATTAAGAGACCAGTTGTGACGATTTGTATTCGGGT 834
298 HisGluThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLe 314
835 TACGAAACTGTCTCTACGACCTCAATGATGGCTCTTAAGTACCTTCATGA 884
314 uAsnThrGluValValGlnLysValArgGluGlu.....ValGlnG 328
885 TCACCCAAAGCTCTTCAAGAACTAAGAGCTGAGCATTTGGCATTCAGGG 934
328 LuLysValGluMetGlyMetTyrThrProGlyLysGlyLeuSerMetGlu 344
935 AAAGAAACGACAG.....GACGAACCACTCGGTCTTGAG 969
345 LeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGluThrLeuAr 361
970 GACGTGAAGTCAATGAAGTTCACCTGAGCTGTGATTATGAGACATCAAG 1019
361 gTLeAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheG 378

1020 ATTGGCAACGATCGTTAATGGGGTCTTAAGCAAAACTACTCGTGACTTGG 1069
378 lLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIle 394
1070 AAATCAACGGTTATTTAATCCCAAAAGATGGAGATTATGATATACAG 1119
395 CysAspThrHisAspValAlaAspValPheProAsnLysGluGluPheG 411
1120 AGGGAATTAATTACGATGCAAAATCTTATGAAGACCATTGATCTTTAA 1169
411 nProGluArgPheMetSerLysGlyLeuGluAspGlySerArgPheAsn 428
1170 TCCATGGAGATGGATGAAGAGAGCTTGGAG.....TCACAAACTCAT 1213
428 yTrIleProPheGlyGlyGlySerArgMetCysValGlyLysGluPheAla 444
1214 GCTTTGTGTTGGAGGTGGGACACAGGCTTGTCTCGTTAAGGAAGTAGGG 1263
445 LysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAsnTr 461
1264 ATTGTGAGATCTCGAGCTTCTCCATTACTTGTGTACGAGATACAGATG 1313
461 pIleLeuSerAsnGlyProProThrMetLysThrGlyProThrIleTyrP 478
1314 GGAGGAATAGAGGGGATGAATTAAATG.....GTGTTTC 1348
478 ro 478
1349 CG 1350
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC35582
seq_documentation_block:
ID AAC35582 standard; DNA; 1667 BP.
XX AAC35582;
AC
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10705.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality:	475.50	Length:	483
Ratio:	1.628	Gaps:	15
Percent Similarity:	60.455	Percent Identity:	27.122

alignment_block:

US-09-668-482-2 x AAC35582 ..

Align seg 1/1 to: AAC35582 from: 1 to: 1667

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
:::||||| :::::||||| ||||| ::::|
109 TTGGGCTCTTCTTGATCATGTGTTGT..TTATGTACTGCTCTCTC.. 153
17 oValLeuLeuPheLeuAlaIaValLysLeuTrpGluMetLeuMetIleA 34
154CGATGGAACCGATGCGATAT. 174
34 rgArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
175TCTAAGAAAGTCTTCCCTCGAACCATGGGC 207
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
|||:::||||| :::::||||| ::::|
208 TGGCCAATATTTGGTGAACCACTGAGTTTCTTAACAAGACCAAGATTTC 257
67 eleuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
|:::| ::::| ::::| ::::| ::::|
258 CATGAAAAAACCAAGACTAAGATATGGGAGTTTCTTCAAGCTTCACATTTC 307
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
::||| |||||::: ||| |||||::: ||| |||
308 TTGGCTGCCCAACAATAGTCTCAATGGACGACAGATTAAACAGATACATT 357
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||:::||||| ||| ::::| ::::| ||::|
358 CTAATGAATGAATCGAAAGGACTTGTTCGGGTTACCGCAATCTATGCT 407
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||||::: :::::||||| ::::|
408 TGATATTCTAGGGACATGCAACATAGCTGCGGTTCAATGCGCCGAGCCACC 457
134 ysAsnLysLysAlaIleMetArgAlaPheSerArgAspAlaLeu... 149
:::| ::::| ::::| ::::|
458 GGCTAATGAGAGCTCGTTGCTTCTTAATAAGCCCAACCATGATGAAA 507
150 GluHisTyrIleProValIleGlnGlnGluValLysSeraIaIleGlnGl 166
:::| ::::| |||:::| :::::| ::::|
508 GACCATCTCTTGCTTAAGATTGATGATTTCATGAGAAACTATCTTTGCGG 557
166 uTrpLeuGlnLysAspSerCysValLeuValTyrProGluMetLysLysL 183
||| ::::| ::::| ||| ::::|
558 TTGGGATGATCTTGAGACA..GTTGATATCCAGAAAAAGACCAACATA 604
183 eumetPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGln 199
::| ||| ::::| ||||| ::::|
605 TGGCATTT.....TTATCATCGTTGTACAAATAGCTGAGACTTTG 645

200 IleLysThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleLy 216
||| ::::| ||| :::::|
646 AAAAAACAGAGTTGAAGAATATAGAACAGAGTTTTCAGCTGTGTGT 695
216 sasLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg 233
:::| ::::| ::::| ::::| ::::|
696 GGGAACCTATCGGTCCCGATCGATATCCCGGAACGAATTACCGCAGTG 745
233 lLeuArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArg 249
|||:::| ||||| ||| :::::|
746 GAGTCCAAGCAGAAACAACATC.....GATAGGTTATTG 780
250 LysLysIleGlnAspAspAsnGluAsnGluGlnLysTyrLysAspAl 266
:::| ::::| ::::| ::::|
781 ACAGAACTGATGCCAAGAAAGAAAGAGTCTGGAGAAACTTTCACAGACAT 830
266 aLeuGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerL 283
||| ||||| ::::| ::::|
831 GTTGGGTTACTTGATG.....AAGAAGCAAGATAACCGATACTTGT 871
283 eu.....GlnAlaMetLysGluAlaIaIaThrGluLeuLeuPheGlyGly 297
||| ::::| ::::| ::::| ::::|
872 TAACCGATTAAGAGATTAAGAGATCAAGTGTAAACGATCTGTGATTCGGGT 921
298 HisGluThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLe 314
:::| |||||:::| ::::| ::::|
922 TATGAGACTGTCTCTACAACTCCATGATGCGCTTAAGTATCTCCATGA 971
314 uAsnThrGluValValGlnLysValArgGluGlu.....ValGlnG 328
:::| ::::| ::::| ::::|
972 TCATCCAAAGCTCTTGAAGAACTCAAGAAACATTTGGCTATAAAGG 1021
1022 AGAGAAACGA.....CCTGACGAACCGCTCCTCCTCGAC 1056
345 LeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGluThrLeuAr 361
:::| ::::| ::::| ::::|
1057 GATATTAAATCGATGAATTCACCTCGAGCTGTGATCTTTGAGACATCAAG 1106
361 gIleAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheG 378
|||:::| ||| ||| ::::|
1107 ATGGCAACGATTGTTAATGGTGTCTTAGGAAACTACTCAGCACTTAG 1156
378 lLeuAsnGlyTyrGlnIleProLysGlyTrpAsnVal...IleTyrSer 393
||||||| ||||| ||||| |||||
1157 AACTCAACGGTTATTTAATCCCAAAAGTTGAGAAATTACGTATACACA 1206
394 IleCysAspThrHisAspValAlaAspValPheProAsnLysGluGluPh 410
:::| ::::| ::::| ::::|
1207 AGAGAGATTAACTATGATACATCT..CTTATGAAGATCCCAATGATCTT 1253
410 eGlnProGluArgPheMetSerLysGlyLeuGluAspGlySerArgPheA 427
|:::| |||:::| ::::| ::::|
1254 TAAACCATGAGATGATGGAAGAAAGAGCTTAGAATCAAGAAGCTATTTC. 1302
427 snTyrIleProPheGlyGlySerArgMetCysValGlyLysGluPhe 443
:::| ||||| ||||| |||||
1303TACTCTTTGGAGGTGAGTTAGGCTTGCCCTGGAAGAAGACTA 1347
444 AlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAs 460
:::| ::::| ::::| ::::|
1348 GGAATCTCGGAAGTCTCAAGCTTCTTCACTACTTTGTTACAAATATATAG 1397
460 nTrpIleLeuSerAsnGlyProProThrMetLysThrGlyProThrIle 476
||| ::::| ::::| ::::|
1398 ATGG...GAAGAGATGAGAGAAAGACAATTAATGCTCTTCCAAAGAGTT 1443

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC39029
seq_documentation_block:
ID AAC39029 standard; DNA; 1685 BP.

XX AAC39029;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 23101.
DE Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EPI033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR	25-AUG-1999;	99US-0150566
PR	26-AUG-1999;	99US-0150884
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154039
PR	16-SEP-1999;	99US-0154408
PR	20-SEP-1999;	99US-01544779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156458
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	08-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0159293
PR	13-OCT-1999;	99US-0159294
PR	13-OCT-1999;	99US-0159295
PR	14-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159638
PR	18-OCT-1999;	99US-0159584
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160815
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989
PR	25-OCT-1999;	99US-0161048
PR	25-OCT-1999;	99US-0161405
PR	25-OCT-1999;	99US-0161405
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161360
PR	26-OCT-1999;	99US-0161361
PR	28-OCT-1999;	99US-0161920
PR	28-OCT-1999;	99US-0161992
PR	28-OCT-1999;	99US-0161993
PR	29-OCT-1999;	99US-0162142

```

alignment_scores:
    Quality: 470.00
    Ratio: 1.632
    Percent Similarity: 59.627
    Length: 483
    Gaps: 9
    Percent Identity: 26.915

```

alignment_block:

US-09-668-482-2 X AAC39029

Align seg 1/1 to: AAC39029 from: 1 to: 1685

14 ILeValLeuProValLeuLeuPheLeuAlaAlaValLysLeuTrpGlume 30
:::||||:||||:||||:||||: |||
139 TTGGTTGTACCAATCCTCATCTATGCTTG.....CTTTTGGAAG 179
30 tLeuMetIleArgValAspProAsnCysArgSerProLeuProProg 47
:::||||: ||| |||:: ||||| |||||
180 AGTGATTGTTCAAGAAGAAAAAGAACAGTAGAGGTAAGCTTCCCTCTG 229
47 lYThrmctGlyLeuProPheIleGlyGluThrLeuGlnLeuIleuGln 63

230 GTTCCATGGGATGGCCTTACTTAGGAGAGACTCTACAACCTCTATTACAA 279

230 GTTCATGGGATGGCTTACTTAGGAGAGACTCTACAACCTATTCCAA 279

64 Arg...ArgLysPheLeuArgMetLysArgGlnLysTyrGlyCysIleTyr 79

280 AACCCCATGTTTCTTCACCTCCAAAGCAAAAGAGATATGGAGAGATATT 329

79 rlysthrhrisleupheglyasnprothrvalargvalmetglyalaaspra 96

330 CAAACCCGATTCCTCGGCTATCCATGCGTGATGTTGGCTAGCCCTGAGG 379

```

50 snvalargintleuleuglghlslsleuvalservalgintyp 112
:::| | |::: :: ::
:::| | |::: :: ::

```

CCGCGGGGCTTGTACCTCAGGCCCAATGTTCAAAACAATTAT 429

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

二
 三
 四
 五
 六
 七
 八
 九
 十
 十一
 十二
 十三
 十四
 十五
 十六
 十七
 十八
 十九
 二十
 二十一
 二十二
 二十三
 二十四
 二十五
 二十六
 二十七
 二十八
 二十九
 三十
 三十一
 三十二
 三十三
 三十四
 三十五
 三十六
 三十七
 三十八
 三十九
 四十
 四十一
 四十二
 四十三
 四十四
 四十五
 四十六
 四十七
 四十八
 四十九
 五十
 五十一
 五十二
 五十三
 五十四
 五十五
 五十六
 五十七
 五十八
 五十九
 六十
 六十一
 六十二
 六十三
 六十四
 六十五
 六十六
 六十七
 六十八
 六十九
 七十
 七十一
 七十二
 七十三
 七十四
 七十五
 七十六
 七十七
 七十八
 七十九
 八十
 八十一
 八十二
 八十三
 八十四
 八十五
 八十六
 八十七
 八十八
 八十九
 九十
 九十一
 九十二
 九十三
 九十四
 九十五
 九十六
 九十七
 九十八
 九十九
 一百

146 refactored with `enum` to

[illegible]

163 Alajeg]UTrl.aug]nlvsasnSercvsa]Iauva]TuvBrcc] 170

580 TCCTTACATCTTGGCCCATATG . CCGATTGTTCTCCACCTACCAAGG 636

179 uMetLysLysLeuMetPheArqIleAlaMetArqIleLeuLeuGlpPheG 196

627 GATGAAGAGTTGCGCTTTGATGTGGGATATTCTAGCCATATTGA... 672

196 LuProgluInIleLysThraspgluIngluLeuValGluAlaPheGlu 212

673CATTTGGAGACTTCTTACAA 693

213 Glumetilelysasleuphe.....Serle 221

694 GAGATCTTGAACATTAACATATTTGTGGACCAAGGCTACAACCTTT 743

221 ucflvlietaspvalpripfneserlglyleuuyrargglyleuargalaarga 238

CCCCCAAGGCTCCCGGACCTCTATCAGAAAGCTCTCAGGCGAGAA / 93

[illegible]

C

Category	Item	Value
A	1	1
	2	1
	3	1
	4	1
	5	1
	6	1
	7	1
	8	1
	9	1
	10	1
B	1	1
	2	1
	3	1
	4	1
	5	1
	6	1
	7	1
	8	1
	9	1
	10	1
C	1	1
	2	1
	3	1
	4	1
	5	1
	6	1
	7	1
	8	1
	9	1
	10	1
D	1	1
	2	1
	3	1
	4	1
	5	1
	6	1
	7	1
	8	1
	9	1
	10	1
E	1	1
	2	1
	3	1
	4	1
	5	1
	6	1
	7	1
	8	1
	9	1
	10	1

[illegible][illegible]

288 [uA]a[atbrc]u[eu]a[phcg]vc]vhi[sc]u[thrtb]v]a[cc]w[th~ 304

ACACATCATCGAGTCTTTTGGCGACACGACACCTACTTCG 878

305 AlaThrSerLeuValmetPheLeuGlyValLeuAsnThrGluValGlyIleVal 321

979 TTAAGTTGATCTTAACTTACATGATGATCAGAACTTCTAGAAGC 1028

3321 sValarglucIuValgIngluLysvalGluMetGlyMetTyrThrProG 338

029 TGTAGGCTGAG...CAAAAGCTATATGACGAAACAGTAGAGAGA 1075

3338 LYLysGlyLeuSerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCys 354

OM of: US-09-668-482-2 to: Issued_Patents_NA:* out_format : pfs
Date: Nov 6, 2001 2:29 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+g2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09668482/runat_05112001_133353_10897/app_query.fasta_1.1680
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09668482@cgn1_1_104 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:
Query: US-09-668-482-2
Query length: 492
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 149.780000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	..
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-3				2563.00	5532.50	1.6e-300 1850
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-3				2563.00	5532.50	1.6e-300 1850
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-5				1738.00	3740.44	1.1e-200 1494
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-5				1738.00	3740.44	1.1e-200 1494
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-31				1734.00	3729.89	4.2e-200 1725
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-38				1169.00	2489.45	5.2e-131 4164
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-36				792.00	1674.93	1.2e-85 2677
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-622-166A-1				552.50	1160.42	5.5e-57 1608
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-11				458.00	974.32	1.3e-46 351
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-11				458.00	974.32	1.3e-46 351
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-991-677-3				343.00	702.63	1.7e-31 1883
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-194-981E-3				339.50	697.18	3.5e-31 1591
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-201-118-10				339.50	695.23	4.5e-31 1852
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-238-821B-10				339.50	695.23	4.5e-31 1852
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05744-10				339.50	695.23	4.5e-31 1852
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-201-118-4				339.50	695.22	4.5e-31 1854
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-238-821B-4				339.50	695.22	4.5e-31 1854
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05744-4				339.50	695.22	4.5e-31 1854
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-091-432-1				330.50	675.75	5.5e-30 1838
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-194-981E-4				328.00	673.62	7.2e-30 1419
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2				320.50	554.37	3.1e-23 4403765
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1				320.50	554.35	3.2e-23 4411529
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-201-118-2				311.50	635.07	1.0e-27 1746
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-238-821B-2				311.50	635.07	1.0e-27 1746
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05744-2				311.50	635.07	1.0e-27 1746
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2				303.00	516.30	4.1e-21 4403765
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1				303.00	516.28	4.2e-21 4411529
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-3				302.00	614.81	1.4e-26 1691
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-313-075A-49				296.00	601.27	7.7e-26 1757
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-201-118-8				293.50	595.32	1.6e-25 1829
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-238-821B-8				293.50	595.32	1.6e-25 1829
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05744-8				293.50	595.32	1.6e-25 1829
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-115-052-1				289.50	589.08	3.7e-25 1509
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-033-055A-1				289.50	587.39	4.6e-25 1722
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-313-075A-29				282.50	572.59	3.0e-24 1665
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-292-768-67				275.50	557.36	2.1e-23 1665
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-881-784-8				274.50	555.19	2.8e-23 1665
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-292-768-3				274.50	555.19	2.8e-23 1665
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-292-768-69				271.50	548.66	6.5e-23 1665
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-201-118-14				270.50	544.85	1.1e-22 1892

/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-238-821B-14 + 270.50 544.85 1.1e-22 1892
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05744-14 + 270.50 544.85 1.1e-22 1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-750-703-1 + 270.00 544.83 1.1e-22 1740
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-11 + 269.50 544.37 1.1e-22 1657
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-13 + 268.50 540.96 1.8e-22 1824

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-3

seq_documentation_block:
; Sequence 3, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1850 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-724-466B-3

alignment_scores:
Quality: 2563.00 Length: 492
Ratio: 5.209 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-668-482-2 x US-08-724-466B-3 ..

Align seg 1/1 to: US-08-724-466B-3 from: 1 to: 1850

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||
88 ATGGGGCTGTACACCCCTTATGTGTCACCTTCTGACACCATCGTGCTACC 137
17 oValLeuLeuPheLeuAlaAlaValIysLeuTrpGluMetLeuMetIleA 34
|||||
138 CGTTTACTCTTCTTCGCCGCGGTGAAGTTGTGGAGATGTTATGATCC 187
34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
|||||
188 GACGAGTCGATCCGAAGTGCAGAAAGTCTTACCGCCAGGATGATGGGC 237
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
|||||

```
238 TTGCCGTTCAATTGGAGAAACGCTCCAGCTGATCCTCCAGAGAGAAAGT 287
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeu 84
|||||
288 TCTGCGCATGAAACGGCAGAAATACGGGTGATCTACAGACGACCTCT 337
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|||||
338 TCGGGAACCCGACTGTGAGGGTGATGGGAGCTGATAATGTGAGGCAGATT 387
101 LeuLeuGlyLuhHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||
388 CTGCTGGCGAACAACAAGCTGGTGTCTGTTCAGTGGCCAGCATCATGAG 437
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||
438 AACCATCCTGGGCTCTGACACCCCTCTCCAATGTCATGGAGTTCAACACA 487
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
|||||
488 AAAACAGAAAAAGGCCATTATGAGGGCGTTCTCTCGAGATGCTCGAGAG 537
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGluTr 167
|||||
538 CACTACATTCCTCCGTGATCCAGCAGAGAGTGAAGAGCGCATACAGAAATG 587
167 pLeuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeu 184
|||||
588 GCTGCAAAAAGACTCCTGCGTGTGTTTATCCAGAAATGAAGAAACTCA 637
184 etPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
|||||
638 TGTTCGGATAGCTATGAGAAATCCTGCTTGGTTTGAACCAAGCAATA 687
201 LysThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleLysAs 217
|||||
688 AAGACGGACGAGCAAGAAGCTGTGGAAGCTTTGAGGAATGATCAAAAA 737
217 nLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyL 234
|||||
738 CTGTCTCTCCTTGCATCGACGCTTCTTTCAGTGGTCTGTACAGGGGTT 787
234 euArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArgLys 250
|||||
788 TGAGGGCACGCAATTCATCTACTCCAAAAATTGAGGAAAAATCATCAGGAAG 837
251 LysIleGlnAspAspAsnGluAsnGluGlnLysTyrLysAspAlaIle 267
|||||
838 AAAATTCAAGATGACGACAAATGAAAAAGAACAGAAATCAAAAGACGCCCT 887
267 uGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerLeuG 284
|||||
888 TCAGCTGTTGATCGAGAACAGCAAGAAAGTACGAACTTTTAAAGTTTGC 937
284 lnaIleMetLysGlnAlaAlaThrGluLeuLeuPheGlyGlyHisGluThr 300
|||||
938 AGCGGATGAAAGAGAGCTACAGAGCTTCTATTGAGAGGTCAATGAACC 987
301 ThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGl 317
|||||
988 ACCGCCAGCAGCTGCAACCTCACTTGTATGTTCTGGGCTCTGAACACAGA 1037
317 uValValGlnLysValArgGluGluValGlnGluLysValGluMetGlyM 334
|||||
1038 AGTGTGCAGAGGTCAAGAGAGAGGTTCAAGAGAAAGTTGAATGGGCA 1087
334 etTyrThrProGlyLysGlyLeuSerMetGluLeuLeuAspGlnLeuLys 350
|||||
1088 TGTATACACCTGGAAGGCTTGAGTATGAGCTGTTGACACAGCTGAAG 1137
351 TyrThrGlyCysValIleLysGluThrLeuArgIleAsnProProValPr 367
|||||
1138 TACACTGGATGTGTGATTAAAGAGACTCTTAGAATCAACCCCTCTGTTCC 1187
```

```
367 oGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnI 384
|||||
1188 CGGAGGATTCAGAGTCGCACACTCAAAAACCTTTGAATTGAATGTTACCAA 1237
384 leProLysGlyTrpAsnValIleTyrSerIleCysAspThrHisAspVal 400
|||||
1238 TTCCTAAAGATGGAACGTCAATTACAGCATCTGTGACACGACGATGTG 1287
401 AlaAspValPheProAsnLysGluGluPheGlnProGluArgPheMetSe 417
|||||
1288 GCCAGCTCTTCCAAACAAGAGAGAGTTCCAGCCGGAGAGATTCAITGAG 1337
417 rLysGlyLeuGluAspGlySerArgPheAsnTyrIleProPheGlyGlyG 434
|||||
1338 CAAAGGCTTGAGAGACGGGCTCCAGGTTTAACTACATCCCTTCGAGAG 1387
434 LysArgMetCysValGlyLysGluPheAlaLysValLeuLeuLysIle 450
|||||
1388 GATCCAGGATGTGTGGGCAAGAGAGTTGCCAAAGTGTACTCAAGATC 1437
451 PheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyPr 467
|||||
1438 TTTTAGTTGAGTTAAACGACGATTCGAAATTGGAATTCCTCAAAACGGACC 1487
467 oProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuProT 484
|||||
1488 CCCGACAAATGAAAAACAGCCGACCTATTACCCAGTGAGACAATCTCCCTA 1537
484 hrLysPheThrSerTyrValArgAsn 492
|||||
1538 CCAAAATTCACCTAGTTATGTGAGAAAT 1563

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-3
seq_documentation_block:
; Sequence 3, Application US/08882164D
; Patent No. 6306624
;
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
;
; ZIP: M5L 1A9
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1850 base pairs
```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-3

alignment_scores: Length: 492
 Quality: 2563.00 Gaps: 0
 Ratio: 5.209
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-668-482-2 x US-08-882-164D-3 ..

Align seg 1/1 to: US-08-882-164D-3 from: 1 to: 1850

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||
88 ATGGGGCTGTACACCCCTTATGGTCACCTTCTCTGCACCATCGTGTACC 137
17 oValLeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
|||||
138 CGTTTACTCTTTCTCGCCGGGTGAAGTTGTGGAGATGTTAATGATCC 187
34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
|||||
188 GACGAGTCGATCCGAACCTGCAGAACTCCTTACCGCCAGGTACCATGGGC 237
51 LeuProPheIleGlyGluThrLeuGluLeuIleLeuGluArgArgLysPh 67
|||||
238 TTGCCGTTCATGGAGAAACGCTCCAGCTGATCCTCCAGAGAAGAAAGTT 287
67 eleuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
|||||
288 TCTGCCGATGAACGCGAGAAATACGGGTGCATCTACAAGACGACCTCT 337
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|||||
338 TCGGGAACCCGACTGTCAAGGGTGATGGAGCTGATTAATGAGGACGATT 387
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||
388 CTGCTGGCGGAACACAGCTGGTCTCTGTTCACTGCGCCACGATCAGTAG 437
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||
438 AACCATCTGGGCTGTGACACCCTCTCCAATGTCCATGAGTTCACACACA 487
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
|||||
488 AAACAAGAAAAAGGCCATTATGAGGGCGTTCTCTCGAGATGCTCTGGAG 537
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
|||||
538 CACTACATTCCTCCGTGATCCAGCAGAGGTGAAGAGCGCCATACAGGAATG 587
588 GCTGCAAAAAGACTCTCGCTGCTGTTTATCCAGAAATGAAGAAACTCA 637
167 pleuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeuM 184
|||||
638 TGTTTCGGATAGCTATGAGATTCCTGCTTGGTTTGAACACAGACCAATA 687
184 etPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
|||||
688 AAGACGAGCAGCAAGAACTGTTGGAAGCTTTTGAGGAATGATCAAAAA 737
201 LysThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleLysAs 217
|||||
217 nLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyL 234
|||||
738 CTGTGTCCTTGGCCAATCGACGTTCTTTCAGTGTGTCGTACAGGGGTT 787
234 euArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArgLys 250

|||||
788 TGAGGGCACCGAATTTCATTCACCTCCAAAATGAGGAAAAACATCAGGAAG 837
251 LysIleGlnAspAspAsnGluAsnGluGlnLysTyrLysAspAlaIle 267
|||||
838 AAAATTCAAGATGACGACCAATGAAAAACGACAGAAATACAAAGACGCCCT 887
267 uGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerLeuG 284
|||||
888 TCAGCTGTGATCGAGACAGCAGAGAAGTACGACCACTTTAGTTTGC 937
284 InAlaMetLysGluAlaAlaThrGluLeuLeuPheGlyGlyHisGluThr 300
|||||
938 AGCGGATGAAAAGACAGCTTACAGAGCTTCTATTGGAGGTGATGAAGCC 987
301 ThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGl 317
|||||
988 ACCGCCAGCACTGCACCTCACTTGCATGTTCTGGGCTCGAACACAGA 1037
317 uValValGlnLysValArgGluGluValGlnGluLysValGluMetGlyM 334
|||||
1038 AGTGTGCAGAAAGGTCAGAGAGAGGTTCAAGAGAGGTTGAATGGGCA 1087
334 etYrThrProGlyLysGlyLeuSerMetGluLeuLeuAspGlnLeuLys 350
|||||
1088 TGTATACACCTGGAAGGGCTTGAGTATGAGCTGTTGGACCAAGCTGAAG 1137
351 TyrThrGlyCysValIleLysGluThrLeuArgIleAsnProProValPr 367
|||||
1138 TACACTGTGATGTGATTAAAGAACTCTTAGAATCAACCTCTGTTCC 1187
367 ogLysIlePheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnI 384
|||||
1188 CGGAGGATTCAGAGTCGCACTCAAAACCTTGAATTGAATGTTACCAAA 1237
384 leProLysGlyTrpAsnValIleTyrSerIleCysAspThrHisAspVal 400
|||||
1238 TTCCTAAAGATGGAACGTCATTCATACAGCATCTGTGACACGACGATGTG 1287
401 AlaAspValPheProAsnLysGluGluPheGlnProGluArgPheMetSe 417
|||||
1288 GCCGACGCTTTTCCAAACAAGAGAGGTCCACGCCGAGAGATTCATGAG 1337
417 rLysGlyLeuGluAspGlySerArgPheAsnTyrIleProPheGlyGlyG 434
|||||
1338 CAAAGTCTGAGGAGCAGGGGTCAGGTTTAACATCATCCTCCCTCGAGAG 1387
434 LysSerArgMetCysValGlyLysGluPheAlaLysValLeuLeuLysIle 450
|||||
1388 GATCCAGGATGTGTGGGCCAAAGAGTTGCCCAAGTGTACTCAAGATC 1437
451 PheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyPr 467
|||||
1438 TTTTAGTTGAGTTAACGACGATTCGAATTGGAATTCTCTCAACGGAACC 1487
467 oProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuProT 484
|||||
1488 CCCGACAATGAACAAGCCGCGACTATTTTACCCAGTGACAAATCTCCCTA 1537
484 hrLysPheThrSerTyrValArgAsn 492
1538 CCAAATTCACCTAGTTATGTCAGAAAT 1563

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-5
seq_documentation_block:
; Sequence 5, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30

```

CORRESPONDENCE ADDRESS:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
ZIP: M5L 1A9
COUNTRY: Canada
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1494 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-5

```

```

alignment_scores:
Quality: 1738.00      Length: 497
Ratio: 4.080          Gaps: 5
Percent Similarity: 85.714  Percent Identity: 68.008

```

alignment block:

US-09-668-482-2 x US-08-724-466B-5 ..

Align seg 1/1 to: US-08-724-466B-5 from: 1 to: 1494

```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1 ATGGGGCTCCCGCGCTGTCGCGCAGTGCCTCTGACCTTCGTGCTGCC 50
17 oValleuLeuPheLeuAlaIaValLysLeuTrpGluMetLeuMetIleA 34
|::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
51 GCTGCTGCTCTTCCTGCGTCGATCAAGCTCTGGACCTGTA CTGCGTGA 100
34 rGArgValAspProAsnGlySerProLeuProGlyThrMetGly 50
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
101 GCGGGCGGACCGCAGTTGTGCCCTCCATTGCCCCCGGAGCTATGGGC 150
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
151 TTCCCTCTTTTGGGGAACCTTGACAGATGTA CTGCAAGGGAAGTT 200
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
201 CTGCAGATGAAGCGCAGAAATACGCTTCATCTACAAGACGATCTGT 250
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
251 TCGGGCGGCCACCGTACGGGTGATGGCGCGGACAATGTGCGGCGCATC 300
101 LeuLeuGlyGlnHisLysLeuValSerValGlnTrpProAlaSerValAr 117
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
301 TTGCTCGAGACGACCGGCTGTGTCGCTCCACTGCGCACGCGTGGTGG 350
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

351 CACCATTTCTGGATCTGCTGCTCCTCTTAACCTGCACGACTCCTGCACA 400
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
401 AGCAGCGCAAGAGGTGATATGCGGCTTCAGCCGCGAGCAGCAGCTCGAA 450
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
451 TGCTACGTGCGGCTGATTCACCGAGAGAGTGGCGCAGCAGCTGAGCAGTG 500
167 PleuGlnLysAspSerCys.....ValleuValTyrProGluM 180
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
501 GCTG.....AGTCGCGCGAGCGCGGCTCTGCTTACCCCGAGG 541
180 etLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
542 TGAAGCGCTCATGTTCGCAATCGCCATGCCATCCTACTGGGCTGCGAA 591
197 ProGluGlnIleLysThrAsp.....GluGlnGluLeuValGluAl 210
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
592 CCC...CAACTGCGCGCGCGAGCGGAGCTCCGAGCAGCAGCTGTGGAGC 638
210 apheGluGluMetIleLysAsnLeuPheSerLeuProIleAspValProp 227
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
639 CTTCGAGGAATGACCCGCAATCTCTTCGCTGCCATCGACGTGCCCT 688
227 heSerGlyLeuTyrArgGlyLeuArgAlaArgAsnPheIleHisSerLys 243
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
689 TCAGCGGCTGTACCGGGGCGATGAAGCGCGAACTCATTCACGCGCGC 738
244 IleGluGluAsnIleArgLysIleGlnAsp...AspAspAsnGluAs 259
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
739 ATCGAGCAGAACATTCGCGCAAGATCTGCGGCTGCGGCGATCCGAGGC 788
259 nGluGlnLysTyrLysAspAlaLeuGlnLeuIleGluAsnSerArgA 276
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
789 GGGCAGGCGCTGCAAGACGCGCTGCTGTGATCGACACTGCTGGG 838
276 rGSerAspGluProPheSerLeuGlnAlaMetLysGluAlaIaThrGlu 292
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
839 AGAGGGGAGAGCGGCTGACATGACGACACTAAAGCAATCTTCAACCGAA 888
293 LeuLeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerLeuVa 309
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
889 CTCTCTTTGGAGGACAGCAACCAACGCGCAGTGACGCCACATCTCTGAT 938
309 lMetPheLeuGlyLeuAsnThrGluValValGlnLysValArgGluGluV 326
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
939 CACTTACCTGGGCTCTACCCACATGTTCTCCAGAAAGTGGAGAAAGAGC 988
326 aLglnGluLysValGluMetGlyMetTyrThrProGlyLysGlyLeuSer 342
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
989 TGAAGAGTAAAGGTTTACTTTGCAAGACAAATCAAGACAAACAACTTGAC 1038
343 MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGluTh 359
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1039 ATGGAATTTTGGAACTTAATAATACATCGGTGTGTTAATAAGAGAC 1088
359 rLeuArgIleAsnProProValProGlyGlyPheArgValAlaLeuLysT 376
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1089 CTTTCGACTGAATCCCGCACTTCCAGAGGTTTCGGCTTCTCTGAAGA 1138
376 hrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrPAsnValIleTyr 392
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1139 CTTTGAATTAATGATACCAAGATTTCCCAAGGCTGGAATGTTATCTTAC 1188
393 SerIleCysAspThrHisAspValAlaAspValPheProAsnLysGluG 409
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1189 AGTATCTGTATCTATGATGTGGCAGAGATCTTCAACCAAGGAAGA 1238
409 uPheGlnProGluArgPheMetSerLysGlyLeuGluAspGlySerArgp 426
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1239 ATTAACTCTGACCGATTCAGTGTCTCTCACCACAGAGATGCATCCAGGT 1288

```



```
|||||
889 CTCCTCTTTGGAGGACACGAAACACGAGCCAGTCAGCCACATCTGTGAT 938
309 lmetPheLeuGlyLeuAsnThrGluValValGlnLysValArgGluGluV 326
: |||
939 CACTTACCTGGGGCTCTACCCACATGTCTCCAGAAAGTGCAGAGAGAC 988
326 aGlnGluLysValGluMetGlyMetYThrProGlyLysGlyLeuSer 342
: |||
989 TGAAGAGTAAGGTTTACTTTGCAAGAGCAATCAAGACAACAAGATTGGAC 1038
343 MetGluLeuLeuAspGlnLeuLysYThrGlyCysValIleLysGluTh 359
|||||
1039 ATGGAATTTTGGAAACAACATTAAATACATCGGGTGTGTATTAAAGAGAC 1088
359 rLeuArgIleAsnProProValProGlyGlyPheArgValAlaLeuLysT 376
|||||
1089 CCTTCGACTGAAATCCCCAGTTCACAGAGGGTTTCGGGTGCTCTGAAGA 1138
376 hrPheGluLeuAsnGlyTYrGlnIleProLysGlyTrpAsnValIleTYr 392
|||||
1139 CTTTGAATTAAATGATATACAGATTCACCAAGGGCTGGAATGTTATCTAC 1188
393 SerIleCysAspThrHisAspValAlaAspValPheProAsnLysGluI 409
|||||
1189 AGTATCTGTGATCTCATGTGATGTGGCAGAGATCTTCACCAACAAGAGA 1238
409 uPheGlnProGluArgPheMetSerLysGlyLeuGluAspGlySerArgP 426
|||||
1239 ATTTAATCCTGACCCGATTCAGTGTCTCTCACCCAGAGGATGATCCAGGT 1288
426 heAsnTYrIleProPheGlyGlySerArgMetCysValGlyLysGlu 442
|||||
1289 TCAGCTTCATTCATTTTGGAGGAGGCTTAGGAGCTGTGTAGGCAAGAA 1338
443 PheAlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCy 459
|||||
1339 TTTGCAAAAATTCTTCTCAAAATATTATTACAGTGGAGCTGGCCAGGCATTG 1388
459 sAsnTrpIleLeuSerAsnGlyProProThrMetLysThrGlyProThrI 476
|||||
1389 TGACTGGCAGCTTCTAAATGACCTCTCAATGAAAAACAGTCCACCGC 1438
476 lETyrProValAspAsnLeuProThrLysPheThrSerTYr 489
::|
1439 TGTATCTGTGGACAATCTCCCTGCAAGATTCAACCCATTTC 1479
```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:us-08-882-164D-31

seq_documentation_block:

```
; Sequence 31, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-31
```

```
alignment_scores:          Quality: 1734.00          Length: 496
                          Ratio: 4.070              Gaps: 4
Percent Similarity: 85.887      Percent Identity: 67.742
```

alignment_block: US-09-668-482-2 x US-08-882-164D-31 ..

Align seg 1/1 to: US-08-882-164D-31 from: 1 to: 1725

```
1 MetGlyLeuTYrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||
25 ATGGGGCTCCCGCGGCTGCTGGCCAGTGCCTCTGCACCTTGCTGTGCC 74
17 oValLeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
|::|
75 GCTGCTGCTCTTCTCTGGCGGCTCAAGCTCTGGGACCTGTACTGTGTGA 124
34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
|||
125 GCAGCCGCGATCGCAGCTGCGCCCTTCCCTTGCCCCCGGTACCATGGGC 174
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
::|
175 TTCCATTTCTTTGGGAAACAATGCAGATGTGTGCTTCAGCGGAGAAATT 224
67 eLeuArgMetLysArgGlnLysTYrGlyCysIleTYrLysThrHisLeuP 84
|||||
225 TCTGCAGATGAAGCGCAGGAAATACGGCTTCATCTACAAGACCATCTGT 274
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|||||
275 TTGGGGGCCCCACGCGTGGGTGATGGGGCGGATATATGTGGCGGCATC 324
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||
325 TTGCTGGGAGAGACACCGGTGTGCTGCGTGCACCTGGCCCCGCTGGTGC 374
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||
375 CACCATCTTGGGCGCTGCTGCTCTCCAACTGCACGATTCTCTGCACACA 424
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
|||||
425 AGCAGCGAAAGAGGTGATATGACAGGCTTCAGCCGCGAGCACTCCAG 474
151 HisTYrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
|||||
475 TGCTACGTGCTCGTATCGCTGAGGAAGTCAAGACAGTGTCTGAGACAGTG 524
167 pleuGlnLysAspSerCys.....ValLeuValTYrProGluM 180
|||||
525 GCTA.....AGCTGGCGCGAGCGGCGGCTCTGCTGTACCCCGAGG 565
```

180 etLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
:::|||||
566 TGAAGCCCTCATGTTCGGCATCCGCATCGGCATCTGCTGGCTGCCGAG 615
197 ProGluGlnIle....LysThrAspGluGlnGluLeuValGluAlaPhe 211
||| |||||
616 CCGGCTCCAGCGGGCGGGCGGAGAGCAACAGCTCGTGAGGCTTT 665
211 eGluGluMetIleLysAsnLeuPheSerLeuProIleAspValProPheS 228
|||||
666 CGAGAGATGACCCGCAATCTTCTCTCCCATTGACGCTGCCCTTTA 715
228 erGlyLeuTyrArgGlyLeuArgAlaArgAsnPheIleHisSerLysIle 244
|||||
716 GCGGCTGTACCGGGCGGTGAAGCGCGGAACCTTATACACGCGCATC 765
245 GluGluAsnIleArgLysLysIleGlnAsp...AspAsnGluAsnGlu 260
|||||
766 GAGGAGACATTCGCCCAAGATCCGCCGCTTCAGGCTACAGAGCCGGA 815
260 uGlnLysTyrLysAspAlaLeuGlnLeuIleGluAsnSerArgArgS 277
:
816 TGGGGGTTGCAAGACGCGCTGCAGCTCTGATTGAGCAGCTCGGGGAGA 865
277 erAspGluProPheSerLeuGlnAlaMetLysGluAlaAlaThrGluLeu 293
||| :
866 GGGGAGAGAGGCTGGATATGCAGGCACCTAAACAATCGTCAACAGAGCTC 915
294 LeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerLeuValMe 310
|||||
916 CTCTTGGTGTCATGAAACTACAGCCAGCTGCTGCACATCAGTATCAC 965
310 tPheLeuGlyLeuAsnThrGluValValGlnLysValArgGluGluValG 327
:::
966 TTACCTAGAGACTTACCCACATGTCTCCAGAAAGTTCGAGAAGAGATTA 1015
327 InGluLysValGluMetGlyMetTyrThrProGlyLysGlyLeuSerMet 343
:::
1016 AGAGCAAGGGCTTACTTTGCAAGAGCAATCAAGACAACAAGTTAGACATG 1065
344 GluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGluThrLe 360
||| :
1066 GAAACTTTGGAACAGCTTAATACATTTGGGTGTCTCATTAAGAGACCCT 1115
360 uArgIleAsnProProValProGlyGlyPheArgValAlaLeuLysThrP 377
||| :
1116 GCGATTGAATCCTCCGGTCCAGGAGGGTTCCGGGTGCTCTGAAGACTT 1165
377 heGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSer 393
|||||
1166 TTGAGCTGAATGATACAGATCCCCAAAGGCTGGAATGTTATTACAGT 1215
394 IleCysAspThrHisAspValAlaAspValPheProAsnLysGluGluPhe 410
|||||
1216 ATCTGTGACACCCAGATGTGGCAGATATCTTCACTTAACAAGAGGAATT 1265
410 eGlnProGluArgPheMetSerLysGlyLeuGluAspGlySerArgPhea 427
|:|
1266 TAAATCCGACCGCTTTTATGTCCTCATCCAGAGATGCTTCCCGGTCA 1315
427 snTyrIleProPheGlyGlySerArgMetCysValGlyLysGluPhe 443
:::
1316 GCTTCATTCATTTGAGAGAGGCTTCGAGCTGTGTAGGCAAGAGTTT 1365
444 AlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAs 460
|:|
1366 GCAAAATTTCTTTAAGATATTACAGTGGAGCTGGCTAGGCACTGTGA 1415
460 nTriPLeuSerAsnGlyProProThrMetLysThrGlyProThrIleT 477
:|
1416 TTGGCAGCTTCTAAATGAGACTCTTACAATGAAGACAAAGCCCACTGTGT 1465

477 yrrProValAspAsnLeuProThrLysPheThrSerTyr 489
|||||
1466 ACCCTGTGACAATCTTCCTGCAAGATTCACCTACTTC 1503

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-38

seq_documentation_block:

; Sequence 38, Application US/08882164D
; Patent No. 6306624

; GENERAL INFORMATION:

; APPLICANT: Petkovich, P. Martin, White, Jay A.,

; APPLICANT: Beckett, Barbara R., Jones, Glenville

; TITLE OF INVENTION: Retinoid Metabolizing Protein

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Blake, Cassels & Graydon

; STREET: Box 25, Commerce Court West

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5L 1A9

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage

; COMPUTER: COMPAQ, IBM PC compatible

; OPERATING SYSTEM: MS-DOS 5.1

; SOFTWARE: WORD PERFECT

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D

; FILING DATE: June 25, 1997

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546

; FILING DATE: June 21, 1996

; APPLICATION NUMBER: 08/724,466

; FILING DATE: October 1, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Hunt, John C.

; REGISTRATION NUMBER: 36,424

; REFERENCE/DOCKET NUMBER: 50767/00010

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 863-4344

; TELEFAX: (416) 863-2653

; INFORMATION FOR SEQ ID NO: 38:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4164 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-882-164D-38

alignment_scores: Length: 1055
 Quality: 1169.00 Gaps: 16
 Ratio: 3.005
Percent Similarity: 36.872 Percent Identity: 29.858

alignment_block:
US-09-668-482-2 x US-08-882-164D-38 ..

Align seg 1/1 to: US-08-882-164D-38 from: 1 to: 4164

34 ArgArgValAspProAsnCysArgSerProLeuProPro..... 46
||| ||||| :|: ||||| |||||
25 AGAGGGGTAGACAGACCAAGATGCTCCCGGCGCCCTCATGCGGGTTG 74
47 .GlyThrMetGlyLeuProPheIleGlyGluThrLeuGlnLeuLeuG 63
||| |||
75 CGGTCTCTCTCTCCACCT.....CCCTCTC 100
63 InArgArgLysPheLeuArgMetLysArgGlnLysTyrGlyCysIleTyr 79
|||||
101 AGCGGAGGAAGTTTCTGCAGATGAAGCGCAGAAATACGGCTTCATCTAC 150
80 LysThrHisLeuPheGlyAsnProThrValArgValMetGlyAlaAspAs 96

330 330
1992 TCCACCTCGTAATTAAAAAGGAGTTGTTGACTGGGCTCTCTTG 2041
330 330
2042 GGCAGAAATTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 2091
330 330
2092 TCCACCACCACCATCACCACCTTTATAGCAAGGTTCTCTTCC 2141
330 330
2142 CTGACCAAGACATGAATAATGTGATAGACCAATAGCTGATCAGGTC 2191
331glu..MetGlyMetTyrThrProGlyLys.. 339
2192 GCAGTGTGTGAGGGCTCAGGGTATGACCTTTATATACCTGATAGCA 2241
339 339
2242 ACATGTCTGGATAATGGGTTAGGCTGAGAGAAGTGGAAGAAGGCC 2291
339 339
2292 ATCAGGCCATCAGCTCTTTCCCTTTATCTCTCTCCATCCAGACGCTTC 2341
339 339
2342 AGGTTAGTTAACAGGTAGTCTCTGCTGGGCTGACTTTTGTGGAGTG 2391
339 339
2392 CCCAGGGATCCATCACTCACTTTTATCTGTTTCCATAGGGCTTACTTT 2441
340GlyLeuSerMetGluLeuLeuAspGlnLeu 349
2442 GCAAGAGCAATCAAGACAAGATTAGACATGGAACCTTTGGCAGAGCTT 2491
350 LysTyrThrGlyCysValIleLysGluThrLeuArgIleAsnProProVa 366
2492 AAATACACTGGGTGTTCATTAAGAGAGACCCCTGCGATGAACTCCGGT 2541
366 LProGlyGlyPheArgValAlaLeuLysThrPheGlu..... 378
2542 TCCAGGAGGGTTTCGGGTCTCTGAAGACTTTTGAGCTGAATGTGAGTG 2591
378 378
2592 CACCTCTCTGCCCCCACCACCCCTGCTCCACGTCCACTCTGCTATGC 2641
378 378
2642 TGTGAGCATCAGCTGCCAGAGCAGTGCGTCACTGCCCTTGACAGTGTG 2691
378 378
2692 CTGCTCTCTATGTGACTGGGAACCAATTGCTCTCTCTTAATGCCAT 2741
378 378
2742 CCATGCTAGTAATGACTTTTGTGTGCAAGCTCAGGGCCGGGATTGTC 2791
379LeuAsn..GlyTyrGlnIleProLys 386
2792 AATCTTAGAATTTTCTTTTAAACAGGATACAGATCCCAAG 2841
387 GlyTyrAsnValIleTyrSerIleCysAspThrHisAspValAlaAspVa 403
2842 GGCTGGAATGTATTATTACAGTATCTGTGACACCCACGATGTGCAATAT 2891

403 LpheProAsnLysGluGlnPheGlnProGluArgPheMetSerLysGlyL 420
2892 CTTCACTAACACAGAGAGATTTAATCCCGACCGCTTATAGTGCCTCATC 2941
420 euGluAspGlySerArgPheAsnTyrIleProPheGlyGlySerArg 436
2942 CAGAGGATGCTTCCCGGTTCAGCTTCATTCATTTGAGAGAGCGCTTCGG 2991
437 MetCysValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeuVa 453
2992 AGCTGTGTAGGCAAGAGTTTGCAAAATTTCTTCTTAAGATATTACAGT 3041
453 LgluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyProProThrm 470
3042 GGAGCTGGCTAGGCACTGTGATGGCAGCTTCTAAATGAGACCTCTACAA 3091
470 eLysThrGlyProThrIleTyrProValAspAsnLeuProThrLysPhe 486
3092 TGAAGACAAGCCCACTGTGTACCTGTGTGACAAATCTCCCTGCAAGATT 3141
487 ThrSerTyr 489
3142 ACCCACTTC 3150

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-36

seq_documentation_block:
: Sequence 36, Application US/08882164D
: Patent No. 6306624
: GENERAL INFORMATION:
: APPLICANT: Petkovich, P. Martin, White, Jay A.,
: APPLICANT: Beckett, Barbara R., Jones, Glenville
: TITLE OF INVENTION: Retinoid Metabolizing Protein
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Blake, Cassels & Graydon
: STREET: Box 25, Commerce Court West
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5L 1A9
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
: COMPUTER: COMPAQ, IBM PC compatible
: OPERATING SYSTEM: MS-DOS 5.1
: SOFTWARE: WORD PERFECT
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/882,164D
: FILING DATE: June 25, 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/667,546
: FILING DATE: June 21, 1996
: APPLICATION NUMBER: 08/724,466
: FILING DATE: October 1, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunt, John C.
: REGISTRATION NUMBER: 36,424
: REFERENCE/DOCKET NUMBER: 50767/00010
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 863-4344
: TELEFAX: (416) 863-2653
: INFORMATION FOR SEQ ID NO: 36:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2677 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-882-164D-36

alignment_scores: length: 673
Quality: 792.00 Gaps: 13
Ratio: 2.891

Percent Similarity: 40.713 Percent Identity: 31.947

alignment_block:

US-09-668-482-2 x US-08-882-164D-36 ..

Align seg 1/1 to: US-08-882-164D-36 from: 1 to: 2677

```
1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
   |||||  ::|||:  ::|||:  |||||:  ::|||:
676 ATGGGCTCCCGGCGCTGCTGGCCAGTGGCTCTGCACCTTCGTCTGCC 725

17 ovalLeuLeuPheLeuAlaValLysLeuTrpGluMetLeuMetIleA 34
   |||||  ::|||:  ::|||:  |||||:  ::|||:
726 GCTGCTGCTCTTCTGCTGCTGCATCAAGCTCTGGGACCTGTACTGCTGA 775

34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
   |||  ::|||  |||||  |||||  |||||
776 GCGGCGCGGACCGCAGTTGTGCCCCCTCCCATTTGCCCCCGGACTATSGS 825

51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgLysP 67
   ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
826 TTCCCTCTTTGGGAAACCTTGACATGNTACTNCAGGTAAGGAGGG 875

67 heLeuArg..... 69
   |||
876 TGGGCGGGACAGGCTGCTTCCCCGGAGCCCGCGGCTCTGGGCTTCT 925

69 ..... 69
926 GCTGAAGTGGGGGTAGGCGCCCCGGGAGGCATGCTATTGCGGCTAGAG 975

69 ..... 69
976 CAGGCTGGCGGAGCGCGGCTCCCGGKMYMCSTCAMGCSRCRWK 1025

70 .....MetLysArg 72
   |||||
1026 TMMCTCCGCCCTTCTCCCAMAGCGGARSAAARKYKGMRGATGAAGCGC 1075

73 GlnLysTyrGlyCysIleTyrLysThrHisLeuPheGlyAsnProThrVa 89
   ::|||:  |||||  |||||  |||||  |||||
1076 AGGAATACGGCTTCATCTACAGACGATCTGTTGCGGCGGCCACCGT 1125

89 lArgValMetGlyAlaAspAsnValArgGlnIleLeuGlyGluHisL 106
   |||||  |||||  |||||  |||||  |||||
1126 ACGGTGATGGGCGCGGACAATGTGGCGCATCTTGTGAGAGACAC 1175

106 ysLeuValSerValGlnTrpProAlaSerValArgThrIleLeuGlySer 122
   ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
1176 GGCTGGTGTGCTGCTCACTGGCCAGCGTGGTGCGCACCATTCGGGATCT 1225

123 AspThrLeuSerAsnValHisGlyValGlnHisLysAsnLysLys.. 138
   |||||  ::|||:  ::|||:  ::|||:  ::|||:
1226 GGCTGCTCTCTAACCTGCACGACTCTCGACACAAGCAGCAAGAAGGT 1275

138 ..... 138
1276 GGGGGCAGAGGCGACGGCTGACAGGAGGGGAGCCCAATTATGAGCG 1325

138 ..... 138
1326 GAATTCCGGCTGATGGATGCTAGCGCGGCTAGCAGCTTGAGGTGGGCT 1375

138 ..... 138
1376 AGGACCCCTTGCCAGCTCCAGGTTAGCTTTCCAGCTCGAGAGAGTGCCAT 1425

138 ..... 138
1426 GTGTCTGGCAGAGACTGGGGGTGTCTGGAAGGGAGCGCGGTAGACGAGAG 1475

139 .....AlaIle 140
   ::|||
```

```
1476 GGGCGGATGAGGCTTTTAAACGCTGTCCCCCTCCTCGGGACTCAGTGATT 1525

141 MetArgAlaPheSerArgAspAlaLeuGlnHisTyrIleProValIleGl 157
   |||||  ::|||:  ::|||:  |||||  ::|||:
1526 ATGCGGCTTCAGCCCGGAGGACACTCGAATGCTACGTGCGGTGATCAC 1575

157 nGlnGluValLysSerAlaIleGlnGluTrpLeuGlnLysAspSerCys. 173
   ::|||:  |||:  ::|||:  ::|||:  |||||
1576 CGAGGAAGTGGGCGAGAGCCTGAGCAGTGGCTG.....AGCTGCC 1616

174 .....ValLeuValTyrProGluMetLysLysLeuMetPheArg 186
   ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
1617 GCGAGCGCGGCTCTGCTGCTTACCCCGAGGTGAAGCGCTCATGTTCCGA 1666

187 IleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIleLysThrAs 203
   |||||  |||||  |||||  |||||  |||||
1667 ATGCGCATGCGCATCTACTGGGCTGCGAACC...CAACTGGCGGCGCA 1713

203 P.....GlnGlnGluLeuValGluAlaPheGlnGluMetIleLysA 217
   |||||  ::|||:  ::|||:  ::|||:  ::|||:
1714 CGGGACTCCGACAGCAGACTTGTGAGGCTTCGAGAAATGACCCGCA 1763

217 snLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg... 232
   |||||  |||||  |||||  |||||  |||||
1764 ATCTCTTCTGCTGCCCATCGACGTGCCCTTCAGCGGCTGTACCGGTA 1813

232 ..... 232
1814 AGGCGGCAACGGGCTGCGGACTAGGGCGCGGACCTGGGCGTCTGCT 1863

233 .....GlyLeuArgAlaArgAsnPheIle 240
   |||  ::|||:  ::|||:  ::|||:  ::|||:
1864 CACCGCCCGCGCTCTCTGCGCTCAGGGCATGAAGCGCGGAACCTCATTT 1913

241 HisSerLysIleGlnGluAsnIleArgLysLysIleGlnAsp...AspAs 256
   |||:  ::|||:  ::|||:  ::|||:  |||||
1914 CACGCGCGCATCGAGCAGAACATTCGCCCAAGATCTGCGGGCTGCGGCG 1963

256 pAsnGluAsnGlnGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGluA 273
   ::|||  |||  |||||  |||||  |||||
1964 ATCCGAGGGGGCCAGGGCTGCAAAAGCGCGCTGCAGCTGTTGATCGAGC 2013

273 snSer..... 274
   ::|||
2014 ACTGTTGGAGAGGGGAGAGAGCGGCTGCACATGCAGGTGAGTAGACCTTC 2063

274 ..... 274
2064 AGACGAGCACTGCGGAGTTTGTCCCTGCTTCCAAAGCGCTGTCC 2113

274 ..... 274
2114 TGGGGCCCCCAAAGCGCGCTGGGGCCACGCTTCTGGAGTGGCGGC 2163

275 .....ArgArgSerAspGluPro... 280
   |||:  ::|||:  ::|||:  ::|||:  |||||
2164 CGGCTCAGACTACAGCTATGGAATCCCGAAGGAGCTGAGACACCCGGT 2213

280 ..... 280
2214 CAGGAGAGCTGCGGAAGGGCTGCGGAGGAAACTGGGAGCATCCCTAGC 2263

280 ..... 280
2264 CTTTAMCAGGTTTCAAGGAAAGTTGGAATTTGCAAAATGTTAATAAA 2313

280 ..... 280
2314 GAACCTTGCGATTTTAATAAACTAAGACTTTAAGTCAAGAGATTCCGGT 2363

280 ..... 280
2364 AGRGCGGGGTGCTACTGCGCTTACTGCTCCAGCTGAACATAAGGAGCTT 2413
```

```
280 ..... 280
2414 GCATTTGTTTAAAGATATGCTTCTTGACTTCTGTACAGCAAAACAT 2463
280 ..... 280
2464 TTAGCCCTTCTAGTCTTCCCTCAGAACTCTCAGTTGATTCGAGTAAT 2513
281 .....PheSerLe 283
2514 CCTTCTGTCAAACCGCAGGACGACTGTGAGAATGTGGGTCTCACTCTAT 2563
283 uGln..AlaMetLysGluAlaAlaThrGluLeuLeuPheGlyHisG1 299
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2564 TCTTAGGACTAAGCAATCTTCAACCGAACTCTCTTTGGAGGACACGA 2613
299 uThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnT 316
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2614 AACCAAGCGGACGACGACACACATCTGTGATCACTTACCTGGGCTCTACC 2663
316 hrGluValValGln 320
      ::|||:||||
2664 CACATGTCTCCAG 2677

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-622-166A-1
```

```
seq_documentation_block:
; Sequence 1, Application -us/08622166A
; Patent No. 5952545
;
; GENERAL INFORMATION:
; APPLICANT: KONCZ, CSABA
; APPLICANT: MATHUR, JAIDEEP
; APPLICANT: SZEKERES, MIKLOS
; APPLICANT: ALTMANN, THOMAS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
; TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
; NUMBER OF SEQUENCES: 4
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,166A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 0147-0153P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: A. thaliana
```

```
; IMMEDIATE SOURCE:
; LIBRARY: lambda gt10
; CLONE: C204
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 48..1466
; US-08-622-166A-1

alignment_scores:
      Quality: 552.50      Length: 485
      Ratio: 1.811      Gaps: 14
Percent Similarity: 62.887      Percent Identity: 29.897

alignment_block:
US-09-668-482-2 x US-08-622-166A-1 ..

Align seg 1/1 to: US-08-622-166A-1 from: 1 to: 1608

19 LeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleArgAr 35
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 CTCTCTCTCTCTCTCTCTCCATCGCGCCGCCGCTTCCCTCCTCTACTCCGCCG 115
35 gValAspProAsnGlySerProLeuProProGlyThrMetGlyLeuP 52
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 TACACGTACCGT...CGGATGGGTCTGCTCCGCGGAAGCCTTGCTCTCC 162
52 ropheIleGlyGluThrLeuGlnLeuIle.....LeuGln 63
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 CTCTGATAGAGAGACTTTTACGCTGATCGGAGCTTACAAAACAGAGAAC 212
64 ArgArgLysPheLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLy 80
      ::|||:||||:||||:||||:||||:||||:||||:||||:
213 CTGAGCCTTTCATCGACGAGAGATAGCCCGGTACGGTTCGTTTCAT 262
80 sThrHisLeuPheGlyAsnProThrValArgValMetGlyAlaAspAsnV 97
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 GACGCATCTTTTGGTGAACCGACGATTTTCTCAGCTGACCCGGAACGA 312
97 alArgGlnIleLeuLeuGlyGlnHisLysLeuValSerValGlnTrpPro 113
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 ACCGGTTGTCTTCTCAGAACGAAAGGAGCTTTTGAGTGTCTTATCCCT 362
114 AlaSerValArgThrIleLeuGlySerAspThrLeuSerAsnValHisG1 130
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 GCTTCATTTGTAACCTTTTGGGGAACACTCTCTGCTTCTTATGAAAGG 412
130 yValGlnHisLysAsnLysLysLysAlaIleMetArgAlaPheSerArg 147
      | ||| ||| ||| ::|: |||: |||: |||: |||: |||: |||:
413 TTCTTTGCATAAA...CGTATGCACTCTCTCACCATGAGCTTGTCTAAT 459
147 sPAlaLeu.....GluHisTyrIleProValIleGlnGlnValLys 161
      ::|: |||: |||: |||: |||: |||: |||: |||: |||:
460 CTTCATCATTAAGACACCATCTCATGTGATATTGACCGGTAGTCCGG 509
162 SerAlaIleGlnIleuTrpLeuGlnLysAspSerCysValLeuValTyrPr 178
      ::|: |||: |||: |||: |||: |||: |||: |||: |||:
510 TTATATCTGATCTTGG.....TCTTCTCGTGTCTCTCATGGA 550
178 ogLuMetLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyP 195
      ||| ||| |||: |||: |||: |||: |||: |||: |||: |||:
551 AGAAGCCAAAAGATAAGCTTTGAGCTAACGGTGAAGCAGTTGATGAGCT 600
195 heGluPro.....GluGlnIleLysThrAspGluGlnIleu 207
      |||: |||: |||: |||: |||: |||: |||: |||: |||:
601 TTGATCCAGGGGATGAGTGAGAGTTTAAGAAAGATATCTTCTGTGC 650
208 ValGluAlaPheGluGluMetIleLysAsnLeuPheSerLeuProIleAs 224
      ::|: |||: |||: |||: |||: |||: |||: |||: |||:
651 ATCGAAGGCTTC.....TTCTCTCTTCCCTCC 679
224 pValProPheSerGlyLeuTyrArg..GlyLeuArgAlaArgAsnPheI 240
      ::|: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

680 TCCTC...TTCTCCACCACTTACCGCAAGCCATCCAAGCGCGAGGAAG 726
240 leHisSerLysIleGluGluAsnIleArgLysIleGlnAspAsp 256
::: :::: :::: :::: :::: :::: :::: :::: ::::
727 TGGCGAGCGGTGACGCGTGGTGTGATGAAGAAG...AGGAGAGAGAG 773
257 AsnGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuIleGluAs 273
::: :::: :::: :::: :::: :::: :::: :::: ::::
774 GAAGAAGAGCGGAGAGAGAAGAAGATATGCTTGGCGGCTGCTT... 818
273 nSerArgArgSerAspGluProPheSerLeuGlnAlaMetLysGluAla 290
819GCCGCGATGATGGATTTCCGATGAAGAGATTGTGACTTCT 861
290 laThrGluLeuLeuPheGlyGlyHisGluThrAlaSerThrAlaThr 306
::: :::: :::: :::: :::: :::: :::: :::: ::::
862 TGGTGGCTTACTTGTGCGCGGTATGAACAACCTCCACGATCATGACT 911
307 SerLeuValMetPheLeuGlyLeuAsnThrGluValGlnLysValAr 323
912 CTCGCCGTCAAAATTTCTCACCGAGACTCCTTTAGCTCTTGGCTCAACTCAA 961
323 gGluGluValGlnGluLysValGluMetGlyMetTyrThrProGlyLysG 340
::: :::: :::: :::: :::: :::: :::: :::: ::::
962 GGAAGAG...CATGAAAAGATTAGG...GCAATGAAGAGATTGCTATATA 1005
340 lLeuSerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIle 356
::: :::: :::: :::: :::: :::: :::: :::: ::::
1006 GTCTTGAATGAGTGAATTACAAAGTCAATGCCATTCAACAAATGTGTGTT 1055
357 LysGluThrLeuArgIleAsnProValProGlyGlyPheArgValAl 373
::: :::: :::: :::: :::: :::: :::: :::: ::::
1056 AATGAGACGCTACGAGTGGCTAACATCATCGCGGTGTTTCAGACGTGC 1105
373 aLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsn 390
::: :::: :::: :::: :::: :::: :::: :::: ::::
1106 AATGACGGATGTTGAGATCAAAAGGTTATAAAATCCAAAAGGTTGGAAG 1155
390 alIleTyrSerIleCysAspThrHisAspValAlaAspValPheProAsn 406
::: :::: :::: :::: :::: :::: :::: :::: ::::
1156 TATTCTCATCGTTAGAGCGGTTCATTTAGACCCCAACCACTTCAAAAGAT 1205
407 LysGluGluPheGlnProGluArgPheMetSerLysGlyLeuGluAspG 423
::: :::: :::: :::: :::: :::: :::: :::: ::::
1206 GCTCGCACTTTCACACCTGGAGATGGCAGAGCAACTCGGTAAACGACAG 1255
423 ySerArgPheAsnTyrIleProPheGlyGlyLysSerArgMetCysValG 440
::: :::: :::: :::: :::: :::: :::: :::: ::::
1256 CCCTTCTAATGTGTTCACACCGTTTGGTGAGGGGCCAAGCTATGTCCCG 1305
440 lLysGluPheAlaLysValLeuLeuLysIlePheLeuValGluLeuThr 456
::: :::: :::: :::: :::: :::: :::: :::: ::::
1306 GTTACGAGCTGGCTAGGGTGCACCTCTGTTCCTTCACCGCCTAGTG 1355
457 GlnHisCysAsnTrpIleLeuSerAsnGlyProProThrMetLysThrG 473
::: :::: :::: :::: :::: :::: :::: :::: ::::
1356 ACAGGCTTCAGTTGGTT.....CCTGCAGAGCAAGACAA 1390
473 yProThrIleTyrProValAspAsnLeuProThrLysPheThrSerTyrV 490
::: :::: :::: :::: :::: :::: :::: :::: ::::
1391 GCTGTTTCTTCCAACTACAGAAGACGCAAGAAAGGTAACCCGATCTTCG 1440
490 alArg 491
::: :::: :::: :::: :::: :::: :::: :::: ::::
1441 TGAAG 1445

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-11
seq_documentation_block:
; Sequence 11, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,

;; APPLICANT: Beckett, Barbara R., Jones, Glenville
;; TITLE OF INVENTION: Retinoid Metabolizing Protein
;; NUMBER OF SEQUENCES: 30
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Blake, Cassels & Graydon
;; STREET: Box 25, Commerce Court West
;; CITY: Toronto
;; ZIP: M5L 1A9
;; COUNTRY: Canada
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
;; COMPUTER: COMPAQ, IBM PC compatible
;; OPERATING SYSTEM: MS-DOS 5.1
;; SOFTWARE: WORD PERFECT
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/724,466B
;; FILING DATE: October 1, 1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/667,546
;; FILING DATE: June 21, 1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hunt, John C.
;; REGISTRATION NUMBER: 36,424
;; REFERENCE/DOCKET NUMBER: 50767/00004
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (416) 863-4344
;; TELEFAX: (416) 863-2653
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 351 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-724-466B-11

alignment_scores:
Quality: 458.00 Length: 117
Ratio: 4.404 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 74.359

alignment_block:

US-09-668-482-2 x US-08-724-466B-11 ..

Align seg 1/1 to: US-08-724-466B-11 from: 1 to: 351

292 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerIle 308
|||||
1 GAACCTCTCTTTGAGAGACACGAAACCAAGCCAGCTGCAGCCACATCTCT 50
308 uValMetPheLeuGlyLeuAsnThrGluValValGlnLysValArgLug 325
::: :::: :::: :::: :::: :::: :::: :::: ::::
51 GATCACTTACCTGGGGCTTACCCACATGTTCCTCCAGAAAGTCCGAGAAG 100
325 luValGlnGluLysValGluMetGlyMetTyrThrProGlyLysGlyLeu 341
||: :::: :::: :::: :::: :::: :::: :::: ::::
101 AGCTGAAGAGTAAGGGTTTACTTTGCAAGAGCAATCAAGACAACAAGTTG 150
342 SerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysG 358
::: :::: :::: :::: :::: :::: :::: :::: ::::
151 GACATGGAATTTTGGACCAACTTAATACATCGGCTGTGTTAAGA 200
358 uThrLeuArgIleAsnProProValProGlyGlyPheArgValAlaLeu 375
|||||
201 GACCCCTCGACTGAATCCCAAGTCCAGAGAGGTTTCGGGTGCTCTGA 250
375 yThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIle 391
|||||
251 AGACTTTTGAATTAATGATACCAAGATTCCCAAGGCTGGAATGTATC 300
392 TyrSerIleCysAspThrHisAspValAlaAspValPheProAsnLysG 408
|||||
301 TACAGTATCTGTGATACTCATGATGTGGCAGAGATCTTCAACCAACAAGGA 350


```
81 rHisLeuPheGlyAsnProThrValArgValMetGlyAlaAspAsnValA 98
   ::: ||| ||| |||
298 CAAGATGGGATCTTACACATGATGGCGGTTCCACACACCGACATGGCTC 347
   |||||:::||||::: |||||::: |||||::: |||||
98 rGlnIleLeuLeuGlyLuhHisLysLeuValSerValGlnTrpProAla 114
   |||||:::||||::: |||||::: |||||::: |||||
348 GCCAGTCCCTCAAGTCCAAAGACACATCTCTCGAACCGG...CCAGCC 394
115 SerValArg.....ThrIleLeuGlySerAspThrLeuSerAs 127
   ::::: ||| ::::: |||
395 ACCATAGCCATCAGCTACCTCACCCTATGACCGAGCCGACATGGCCTTCC 444
127 nValHisGly...ValGlnHisLysAsnLysLysLysAlaIleMetArgA 143
   ::::: ||| ::::: ||| ::::: |||
445 TCACACTAGCGCCCGTTTGGCGGTACAGATCGCTAAACTCTCCGTCATGAAAT 494
143 IapHeSerArgAspAlaLeuGlnHisTyrIleProValIleGlnGlnGlu 159
   ||||| ||| ::::: |||
495 TATTTAGCCCGGAAACGAGCCGAGTCGTGG...GAGTCGGTCGAGACGAG 541
160 ValLysSerAlaIleGlnGlnIleTrpLeuGlnLys...AspSerCysVal.. 174
   ||| |||||::: ||| ::::: |||
542 GTCGACTCGGCGAGTACGAGTGGTCGCTCCAAATATGCGGTGCGACGGTGAA 591
175 .....LeuValTyrProGlnMetLysLysLysLeuMetPheArgIleA 188
   |||||::: |||||::: |||||::: |||||
592 TATCGCGGAGCTGGTTTGGCTCTGACGAAAGATATTTACTTACAGGGCGG 641
188 IametArgIleLeuLeuGlyPheGluProGlnIleLysThrAspGlu 204
   ||::: |||
642 CTTTGGGACGATC.....TCGCATGAGGACCAG 670
205 GlnGluLeuValGlnAlaPheGlnGlu.....MetIleLysAsnLeuPh 219
   ::|||::: ||| ::::: |||
671 GACGAGTTCGTGGCCATCTGCAAGAGTTTTCGACGCTGTTGGTCTTT 720
219 eSerLeuProIleAspValProPhe...SerGlyLeuTyrArgGlyLeuA 235
   |::: ||| ::::: ||| ::::: |||
721 TAAATATAGCTGATTTTATCCCTTGGCTCAAAATGGGTTCCTCAGGGGATTA 770
235 rGAlaArg.....AsnPheIleHisSer 242
   ::::: |||
771 ACGTCAGGCTCAACAGGACGAGGGCGCTTGATGGCTTATTGACAAG 820
243 LysIleGlnLuhAsnIleArgLysLysIleGlnAspAspAsnGluAs 259
   |||::: ||| ::::: |||
821 ATCATCGACGATCATATACAGAAGGGAGTAAAACTCGGAGGAGGTTGA 870
259 nGlnGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGlnAsnSerArgA 276
   :::: ||| ::::: |||
871 TACTGATATGGTAGATGATTTACTTGCCTTTTACGCTGAGGAAGCCAAAG 920
276 rGSerAspGluProPheSerLeuGlnAlaMetLys..... 287
   |||::: |||
921 TAAGCGAATCTGACGATCTTCAAAATTCATCAAACTCACCAAGACAAAC 970
288 ...GluAlaAlaThrGlnLeuLeuPheGlyGlyHisGlnThrThrAlaSe 303
   ::||| ::::: ||| |||||::: |||||
971 ATCAAAGCTATCATGACGTAATGTTGGAGGAGCCGAAACGGTGGCGTC 1020
303 rThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGluValValG 320
   |::: ||| ::::: |||
1021 CGCGATTGAATGGGCCATGACGAGGAGTGATGAAAAAGCCAGAAAGATCTAA 1070
320 IuhLysValArgGlnGlnValGlnGlnLysValGlnMetGlyMetTyrThr 336
   ::||| |||||::: ||| ::::: |||
1071 AGAAGGTCCAAAGAAGAACTCGCGGTGGGTGGGCTTGACCGCGGAGTTC 1120
337 ProGlyLysGlyLeuSerMetGlnLeuLeuAspGlnLeuLysTyrThrG1 353
   ||| ::::: ||| |||
1121 GAAGAGAAAGAC.....TTCGAGAAGCTCACCTTACTTGAA 1155
353 yCysValIleLysGlnThrLeuArgIleAsnProProValProGlyGlyP 370
```

```
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1156 ATGCGTACTGAAGGAAGTCTTCCGCTCCACCCACCCATCCACTCC 1205
370 heArgValAlaLeuLysThrPheGlnLeuAsnGlyTyrGlnIleProLys 386
   ::::: ||| ::::: ||| ::::: |||
1206 TCCACGAGACTGCCGAGGAGCGCCGAGGTGCGCGCTACTACATTCGCGG 1255
387 GlyTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaAspVa 403
   ::|||::: |||
1256 AAATCGCGGTGATGATCAACCGCGTCCCATCGGCCGGGACAGAACTC 1305
403 IpheProAsnLysGlnGlnPheGlnProGluArgPheMetSerLysGlyL 420
   ::::: ||| ::::: |||
1306 GTGGGCCGACCCAGATACGTTTAGGCCCTCCAGGTTTCTCAAGACGGTG 1355
420 euGluAsp.....GlySerArgPheAsnTyrIleProPheGlyGlyGly 434
   :: ||| |||||::: |||||::: |||||
1356 TGCCCGATTTCAAAGGGAACAACCTTCGAGTTTCATCCCATTCGGGTCA 1405
435 SerArgMetCysValGlyLysGlnPheAlaLysValLeuLeuLysIlePh 451
   ||| ||| ||| ::::: |||
1406 CGTCGGTCTTGCCCCGCTATGCAACTCGGACTCTACGCGCTGAGACAG 1455
451 eLeuValGlnLeuThrGlnHisCys...AsnTrpIleLeuSerAsnGly 466
   ::::: ||| |||||::: |||
1456 TGtGGCTCACCTCCTT...CACTGTTTCACGTGGAGTTGCCGACGGG 1501

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-194-981E-3
seq_documentation_block:
; Sequence 3, Application US/08194981E
; Patent No. 5886157
; GENERAL INFORMATION:
; APPLICANT: GUENGERICH, F. Peter
; APPLICANT: SANDHU, Zuyu
; APPLICANT: SANDHU, Punam
; APPLICANT: GILLAM, Elizabeth M. J.
; TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF
; TITLE OF INVENTION: HUMAN
; TITLE OF INVENTION: CYTOCHROME P450
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: Suite 1200, 127 Peachtree Street, NE
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,981E
; FILING DATE: February 10, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Selby
; REGISTRATION NUMBER: 38,298
; REFERENCE/DOCKET NUMBER: 22000.0022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1591 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```



```

184 etpHeArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
      ::::::::::::::::::::|||:::
492 CCTCACCCTGTGATCCCACTTTCATCTGGCTGTGCTCCCTGCATGTG 541
201 .....LysThrAspGluGln.. 205
      |||:::
542 ATCTGCTCCATTATTTTCCATAAACGTTTGTGATTATAAAGATCAGCAATT 591
206 ...GluLeuValGluAlaPheGluMetIleLys..... 216
      ::::::::::::::::::::|||:::
592 TCTTAATTAAATGAAAAAGTTGAATGAAAAACATCAAGATTGTGACAGCC 641
217 .....AsnLeuPheSerLeuProIleAspValProPhe 227
      ||| ||||| ||||| |||
642 CCTGGATCCAGATCTGCATTAATTTTCTCTATCATTTGATTAC...TTC 688
228 SerGlyLeuTyrArgGlyLeuArgAlaArgAsn.....PheIleHis 242
      ||| ::::::::::||| ::::||| ||::: ||
689 CCGGAACCTGCACAACAATTAA...CTTAAAAACGTTGCTTTTATGAAAAAG 735
242 rLysIleGluGlnAsnIleArgLysIleGlnAspAspAsnGlu 259
      | ||| ||::: ::::: ::::: |||
736 TTATATTTTGGAAAAAGTAAAGAACCAACCAAGATCAATGACATGAACA 785
259 snGluGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGluAsnSerArg 275
      || ||| :::: ||::: ::::: ::::: |||
786 ACCCTCAGAGACTTATGTATGTCTCTGATGAAAAATGAGAGAAGAAAAAG 835
276 ArgSerAsp...GluProPheSerLeuGlnAlaMetLysGluAlaAlaTh 291
      ::::::::::::::::::::||| ||||| ::::: |||
836 CACAACCAACCATCTGAATTTACTATGAAAAGCTTGAAAAACACTGCAGT 885
291 rGluLeuPheGlyGlyHisGluThrAlaSerThrAlaThrSerL 308
      ::::::::::::::::::::||| ||||| ::::: |||
886 TGACTTGTGTGGAGCTGGAGACAGACAGACAACACCAACCTGAGATATG 935
308 euValMetPheLeuGlyLeuAsnThrGluValValGlnLysValArgL 324
      ::::::::::::::::::::||| ||||| ::::: |||
936 CTCTCCTTCTCTCTGCTGAAGCACCAAGGTCACAGCTAAAGTCCAGAA 985
325 GluValGlnGlnLysValGlnMetGlyMetTyrThrProGlyLysGly 341
      ||| ::::: ||| ::::: |||
986 GAGATTGAACGTGTGATT.....GGCAGAAACCG 1014
341 user.....MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysVal 356
      ||| ||::: ||::: ::::: ||| |||
1015 GAGCCCTGTGATGCAAGACAGAGACCAATGCCCTACACAGATGCTGTG 1064
356 IeLysGluThrLeuArgIleAsnProProValProGlyGlyPheArgVal 372
      :: ||::: ||| ::::: |||
1065 TGCACGAGGTCCAGAGATACATTGACCTTCTCCCAACCAAGCTGCCCAT 1114
373 AlaLeuLys...ThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTr 388
      ||::: ||::: ::::::::::||| ||||| |||||
1115 GCAGTGACCTGTGACATTAAATTTCAGAAACTATCTCATTTCCCAAGGCGAC 1164
388 pasnValIleTyrSerIleCysAspThr...HisAspValAlaAspValP 404
      ::::::::::||| ::::: ||||| ::::: ||
1165 AACCATATTAAATTCCCTGACTTCTGTGCTACATGACACAAAGAA...T 1211
404 heProAsnLysGluGluPheGlnProGluArgPheMetSerLysGlyLeu 420
      ||||| ||| ||| ::::: |||
1212 TTCCCAACCCAGAGATGTTGACCCCTCATCTCACTT.....CTG 1249
421 GluAspGlySerArgPheAsn.....TyrIleProPheGlyGly 433
      ::::::::::||| ::::: ||||| ::::: |||
1250 GATGAAGGTGGCAATTTTAAGAAAGTAATACTTCATGCCCTTTCTCAGC 1299
433 yGlySerArgMetCysValGlyLysGluPheAlaLysValLeuLeuLysI 450
      ::::::::::||| ::::: ||||| ::::: |||
1300 AGGAAAAACGATTTGTGTGGAGAGAGCCCTGGCCCGCATGAGCTGTTT 1349

```

```

450 IePheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGly 466
      ::::::::::::::::::::|||:::
1350 TATTCCTGACCTCCATTTTACAGAACTTTAACTGAAATCTCTGTTGAC 1399
467 ProProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuPr 483
      ||| ::::: ||| ||::: |||
1400 CCAAGAAGACCTTGACACCACTCCAGTTGTGCATGTGATTTGCTCTGTGCC 1449
483 o 483
      |
1450 G 1450

```

seq_name: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq:PCT-US95-05744-10

seq_documentation_block:

; Sequence 10, Application PC/TUS9505744

; GENERAL INFORMATION:

; APPLICANT: GOLDSTEIN, Joyce A.

; APPLICANT: ROMKES-SPARKS, Marjorie

; APPLICANT: DE MORRIS, Sonia M.F.

; TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN

; TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend Khourie and Crew

; STREET: 379 Lytton Avenue

; CITY: Palo Alto

; STATE: California

; COUNTRY: US

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/05744

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/238,821

; FILING DATE: 06-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/201,118

; FILING DATE: 22-FEB-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/864,962

; FILING DATE: 09-APR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Dow, Karen B.

; REGISTRATION NUMBER: 29,684

; REFERENCE/DOCKET NUMBER: 15280-192-1-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 326-2400

; TELEFAX: (415) 326-2422

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1852 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; PCT-US95-05744-10

alignment_scores:

Quality: 339.50

Ratio: 1.183

Percent Similarity: 55.513

Percent Identity: 26.886

alignment_block:

US-09-668-482-2 x PCT-US95-05744-10

..

OM of: US-09-668-482-2 to: EST:* out_format : pfs
Date: Nov 6, 2001 2:26 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+pn.model -DEV=xlp
-O=/cgn2_1/USPRO.spool/US09668482/runat_05112001_133353_10873/app-query.fasta_1.1680
-DB=EST -QFMT=fastap -SUFFIX=1st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.500 -GAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09668482_@CGN1_1_5307 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-668-482-2
Query length: 492
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 2793.150000

score_list:

Sequence	Strd Orig	ZScore	Score Len	Documentation
gb_est43:AW174347	+ 1083.00	2243.53	7.6e-116	720 ! AW174347 f142a06.y1 Sugano Kawa
gb_est23:AI667038	+ 997.00	2064.63	7.0e-106	640 ! AI667038 fc24h03.y1 zebrafish W
gb_est43:AW128823	+ 962.00	1990.82	9.0e-102	662 ! AW128823 fe37g11.y1 zebrafish W
gb_est29:AL532444	- 867.00	1786.23	2.2e-90	1028 ! AL532444 AL532444 LTI_NFL001_N
gb_est29:AL539668	+ 855.00	1761.72	5.2e-89	974 ! AL539668 AL539668 LTI_NFL013_FBI
gb_est17:BE189825	+ 807.00	1665.17	1.2e-83	696 ! BE189825 db61c05.y1 Wellcome CF
gb_est29:AL532445	+ 804.00	1655.84	4.1e-83	893 ! AL532445 AL532445 LTI_NFL001_NH
gb_est22:AI584636	+ 792.00	1637.04	4.6e-82	530 ! AI584636 fb81e05.y1 zebrafish W
gb_est43:AW171465	- 765.50	1576.99	1.0e-78	765 ! AW171465 f142a06.x1 Sugano Kawa
gb_est48:AW513600	+ 739.50	1525.07	7.9e-76	618 ! AW513600 x047b06.x1 NCI_CGAP_Ut
gb_est51:AW765767	+ 722.00	1487.41	9.9e-74	669 ! AW765767 da77a02.y1 Harland sta
gb_est29:AL539667	- 698.00	1434.20	9.1e-71	847 ! AL539667 AL539667 LTI_NFL013_FBI
gb_est101:BG515142	+ 672.00	1384.49	5.3e-68	571 ! BG515142 dad60f10.y1 Wellcome C
gb_est4:AA239785	+ 637.00	1313.37	4.9e-64	474 ! AA239785 mx80a03.r1 Soares mous
gb_est72:BE236243	+ 633.50	1304.51	1.5e-63	537 ! BE236243 143959 MARC lPig Sus s
gb_est102:BG578446	+ 632.00	1300.07	2.7e-63	597 ! BG578446 de99c02.y1 Wellcome CF
gb_est21:AI545454	- 628.00	1288.35	1.2e-62	784 ! AI545454 fb81e05.x1 zebrafish W
gb_est99:BG364937	+ 624.00	1284.93	1.9e-62	522 ! BG364937 dc93f02.y1 NICHHD XGC C
gb_est87:BF385034	+ 619.50	1275.10	6.6e-62	539 ! BF385034 602045632F1 NCI_CGAP_I
gb_est89:BF533769	+ 619.50	1274.55	7.1e-62	564 ! BF533769 602073982F1 NCI_CGAP_I
gb_est43:AW128313	- 611.50	1255.74	7.9e-61	666 ! AW128313 fe37g11.x1 Soares NSF
gb_est82:BF055367	- 611.00	1255.07	8.7e-61	646 ! BF055367 7j78f12.x1 zebrafish W
gb_est82:BF047649	+ 595.50	1225.40	3.9e-59	512 ! BF047649 dc80h02.y1 NICHHD XGC C
gb_est87:BF426259	+ 581.00	1193.84	3.2e-57	563 ! AM595907 da29g11.y2 Xenla 13LiC
gb_est40:AV668503	+ 553.00	1137.30	3.1e-54	471 ! AV668503 AV668503 Sugano Kawa
gb_est22:AI626269	+ 550.00	1130.93	7.1e-54	474 ! AI626269 fc12e02.y1 zebrafish W
gb_est85:BF236872	+ 547.50	1123.98	1.7e-53	545 ! BF236872 602027980F1 NCI_CGAP_I
gb_est40:AV668461	+ 547.00	1124.18	1.7e-53	492 ! AV668461 AV668461 Sugano Kawa
gb_est47:AW460650	- 510.00	1041.76	6.6e-49	731 ! AW460650 da29g11.x1 Xenla 13LiC
gb_est107:R51129	+ 502.50	1030.45	2.8e-48	509 ! R51129 yg71b05.r1 Soares infant
gb_est43:AW199734	+ 481.00	978.15	2.3e-45	919 ! AW199734 da06g04.y1 Xenopus lae
gb_est97:BG161868	- 470.00	958.64	2.8e-44	687 ! BG161868 df69e06.x1 Xenopus lae
gb_est44:AW244588	- 458.00	928.24	1.4e-42	1054 ! AW244588 BR_END09D04ps2 Bain Ra
gb_gss5:CNS05P0U	+ 457.50	938.58	3.7e-43	415 ! AW640733 b198d01.w1 Blackshear/
gb_est51:AW767659	- 456.00	929.57	1.2e-42	671 ! AW767659 da77a02.x1 Harland sta
gb_est14:AI004222	- 448.00	915.62	7.0e-42	532 ! AI004222 ou03e01.x1 Soares tota
gb_est47:AW488377	- 436.50	891.06	1.6e-40	523 ! AW488377 UI-M-BH3-arp-h-01-0-UI
gb_est71:BE188917	- 425.00	865.47	4.3e-39	623 ! BE188917 db61c05.x1 Wellcome CF
gb_gss4:CNS04UDT	- 423.50	857.89	1.1e-38	895 ! AL307658 Tetraodon nigroviridis

gb_est51:AW766054	+ 402.00	816.46	2.3e-36	664 ! AW766054 da81f06.y1 Harland
gb_est101:BG487359	- 388.00	788.69	8.2e-35	583 ! BG487359 dad56d03.x1 Wellcom
gb_est75:BE506442	+ 387.00	791.04	6.1e-35	405 ! BE506442 db83b11.y1 Wellcome
gb_est100:BG408452	+ 385.00	790.03	6.9e-35	312 ! BG408452 dd01g02.y1 NICHHD XG

seq_name: gb_est43:AW174347

seq_documentation_block:

LOCUS AW174347 720 bp mRNA EST 16-NOV-1999
DEFINITION f142a06.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
2640274 5' similar to SW:CP26_BRARE P79739 CYTOCHROME P450 26 ;
mRNA sequence.

ACCESSION AW174347 GI:6440295
VERSION AW174347.1
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio

REFERENCE
AUTHORS Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter
,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
Washu zebrafish EST Project 1999
Unpublished (1999)
Other_ESTs: f142a06.x1
Contact: S.L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

TITLE
JOURNAL
COMMENT
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: T3 ET from Amersham
High quality sequence stop: 478.
Location/Qualifiers
1. 720
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone_11b="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: PME18S-FL3; site_1: DraIII (CACTGTGTG);
site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT];
double stranded cDNA was ligated to a DraIII adaptor
[TGTGGCCCTGTG], digested and cloned into distinct DraIII
sites of the PME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTCTGCTCTAAAGCTGG and 3' end
primer CGACCTGCAAGCTGAGCACA."

FEATURES

source

BASE COUNT	172 a	176 c	196 g	176 t
ORIGIN				
alignment_scores:	Quality: 1083.00	Length: 232		
	Ratio: 4.857	Gaps: 0		
Percent Similarity:	96.121	Percent Identity: 90.948		

alignment_block:

US-09-668-482-2 x AW174347 ..

Align seg 1/1 to: AW174347 from: 1 to: 720

1	MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr	17
25	ATGGGGCTGTACACCCCTTATGGTTCACCGTTGTCTGTTCCATCGTGCACC	74
17	oValLeuLeuPheLeuAlaIaValLysLeuTrpGluMetLeuMetIleA	34
75	CGTTTACTCTTTCCTCGCCGGGTGAAGTTGTGGAGATGTTAATGATCC	124
34	rgArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly	50
125	GACGAGTCGATCCGAACTGCAGAGTCCCTTACCGCCAGGTACCATGGGC	174
51	LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh	67
175	TTGGCGTTCATTGGAGAACGCTCCAGCTGATCCTCCAGAGAAGAAAGTT	224
67	eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP	84
225	TCCTGCCGATGAACCGGACGAATAACGGGTGCATCCTCAAGACGCACCTCT	274
84	heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle	100
275	TCGGGAACCCGACTGTCAAGGTGATGGAGCTGATATGTGAGCAGATT	324
101	IeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr	117
325	CTGCTGGGCGAACAACAAGCTGGTGTCTCTTCAGTGGCCAGCATCAGTGAG	374
117	gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL	134
375	AACCATCTGGGCTCTGACACCCCTCTCCAATGTCCATGGAGTTCACACA	424
134	ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu	150
425	AAAACAAGAAAAAGGCCATTATGAGGGCGTCTCTCGAGATGCTCTGGAG	474
151	HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr	167
475	CACATCATTTCCCGTGTATCCATCAGGAGGTGAAGAGCGCCATACGGAATG	524
167	PleuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeuM	184
525	GCTGCAAAAAGACTCCTGCGCTGCTTATCCAGATATGAGATACTCA	574
184	etPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGlnGlnIle	200
575	TGTTGCGGATAGCTATGAGAAATCCTGCTTGGTGTGAACCAACAATA	624
201	LysThrAspGluGlnIleuValGluAlaPheGluGluMetIleLysAs	217
625	ATGACTGACGAGCATGAGCTGTGGGAAGCTTTGGAGCGCAATGATCAGCGA	674
217	nLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg	232
675	CGTGGCTCTGCTGGCAATCCAGTTCCTTTCATTTGCTGTCTACAGG	720

REFERENCE
AUTHORS

1 (bases 1 to 640)
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
, K., Steptoe, M., Theisling, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.,
and Wilson, R.

TITLE	Washu Zebrafish EST Project 1998
JOURNAL	Unpublished (1998)
COMMENT	Contact: Stephen L. Johnson

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark, cDNA Library Arrayed by:
Matthew Clark, DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
RessourcenZentrumPrimardatenbank, Berlin, Germany (web address:
www.rzpd.de)

FEATURES

Location/Qualifiers
1. .640

BASE COUNT	156 a	162 c	167 g	155 t
ORIGIN				

```
alignment_scores:
  quality: 997.00      Length: 193
  ratio: 5.166         Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.0000
```

alignment_block:
US-09-668-482-2 x AI667038

Align seg 1/1 to: AI667038 from: 1 to: 640

1 MetGlyLeuTyrrThrLeuMetValThrPheLeuCySThrIleValLeuPr 17
 |||||
 62 ATGGGGCTGTACACCCCTTATGGTCACCTTCTGTGCACCATCGTGCTACC 111
 |||||
 17 ovalleuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
 |||||

112 CGTTTACTCTTTCTCCGCCGGGTGAAGTTGTGGAGATGTTAATGATCC 161
34 rgArGValAspProAsnCysArGSerProLeuProProGlyThrMetGly 50
162 GACGAGTCGATCCGAACACTGCAGAACTCCTTACCCGCCAGGTACCATGGCC 211
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArGArgLysPh 67
212 TTGCCGTTCAATGGAGAAACGCTCCAGCTGATCTCTCCAGAGAAGGAGATT 261
67 eleuArGmetLysArGlnLysTyrgLysIleTyrrLysThrHisLeuP 84
262 TCTGCGCATGAACGCGCAGAAATACGGGTGCATCTTACAAGACGACCTCT 311
84 heGLyAsnProThrValArGValMetGLyAlaAspAsnValArGlnIle 100
312 TCGGGAACCCGACGTGCAGGGTGATGGAGCTGATATGTGAGCGAGATT 361
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
362 CTGCTGGGCGAACAAGCTGGTGTCTGTCTCAGTGGCCAGCATCAGTGAG 411
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
412 AACCATCTGGGCTGTGACACCCCTCTCCAATGTCCATGGAGTTCACACA 461
134 ysAsnLysLysLysAlaIleMetArGAlaPheSerArGAspAlaLeuGlu 150
462 AAAACAAGAAAAAGCCATTATGAGGGCGTCTCTCGAGATGCTTGAG 511
151 HistYrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
512 CACTACATTCCTCGTATCCAGCAGAGAGTGAAGAGCCCATACAGGAATG 561
167 pleuGlnLysAspSerCysValLeuValTyrrProGluMetLysLysLeuM 184
562 GCTGCAGAAAAGACTCCTGGGTGCTGTTATCCAGAAATGAAGAACTCA 611
184 etPheArGileAlaMetArGileLeuLeu 193
612 TGTTTCGGATAGCTATGAGAATCCTGCTT 640
seq_name: gb_est43:AW128823
seq_documentation_block:
LOCUS AW128823 662 bp mRNA EST 25-OCT-1999
DEFINITION fe37g11.y1 zebrafish Washu MPIMG EST Danio rerio cDNA 5' similar to
SW:CP26_BRARE P79739 CYTOCHROME P450 26 ;, mRNA sequence.
ACCESSION AW128823
VERSION AW128823.1 GI:6116737
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
REFERENCE
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Riltter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
TITLE Washu zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Other_ESTS: fe37g11.x1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Reissourcenzentrum/PrimaDatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 446.
location/Qualifiers
1..662
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish Washu MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="X11-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'pGACTAGTTCTAGATCGGAGCGGCCCTTTTCTTTTCTTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

BASE COUNT 156 a 174 c 174 g 155 t 3 others
ORIGIN
alignment_scores:
Quality: 962.00 Length: 193
Ratio: 5.090 Gaps: 0
Percent Similarity: 97.927 Percent Identity: 97.409
alignment_block:
US-09-668-482-2 x AW128823 ..
Align seg 1/1 to: AW128823 from: 1 to: 662
1 MetGlyLeuTyrrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||
84 ATGGGGCTGTACACCTTATGTCACCTTCTCTGACACCATCGTGTGCC 133
17 oValLeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
|||||
134 CGTTTACTCTTTCTCGCCCGGTGAAGTTGTGGAGATGTTAATGATCC 183
34 rgArGValAspProAsnCysArGSerProLeuProProGlyThrMetGly 50
|||||
184 GACGAGTCGATCCGAACACTGCAGAACTCCTTACCCGCCAGGTACCATGGCC 233
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArGArgLysPh 67
|||||
234 TTGCCGTTCAATGGAGAAACGCTCCAGCTGATCTCTCCAGAGAAGGAGATT 283
67 eleuArGmetLysArGlnLysTyrgLysIleTyrrLysThrHisLeuP 84
|||||
284 TCTGCGCATGAACGCGCAGAAATACGGGTGCATCTTACAAGACGACCTCT 333
84 heGLyAsnProThrValArGValMetGLyAlaAspAsnValArGlnIle 100
|||||
334 TCGGGAACCCGACGTGCAGGGTGATGGAGCTGATATGTGAGCGAGATT 383

101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
434 AACCATCTGGGCTCTGACACACCCTCTCCAATGTCATGAGTTCAACACA 483
134 ySAsnLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
484 AAAACAAGAAAAAGCCATTATGAGGCGCTTCTCCAGATGCTCGAG 533
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGluTr 167
534 CACTACATTCCTCCGTGATCCAGCAGAGGTGAAGCGCCATACAGGAATG 583
167 pLeuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeuM 184
584 GCTGCAAAAAGACTNCTGCGTGTGCTGGTTATCCAGANATGAAGAACTCA 633
184 etpHeArgIleAlaMetArgIleLeuLeu 193
634 TGTNTCGGATAGCTATGAGAATCCTGCTT 662
seq_name: gb_est29:AL532444

seq_documentation_block:
LOCUS AL532444 1028 bp mRNA EST 13-FEB-2001
DEFINITION AL532444 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSODM014YL13 3
prime, mRNA sequence.
ACCESSION AL532444
VERSION AL532444.1 GI:12795937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1028)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. 1028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODM014YL13"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 265 a 244 c 225 g 282 t 12 others
ORIGIN

alignment_scores:
Quality: 867.00 Length: 270
Ratio: 3.905 Gaps: 3
Percent Similarity: 82.222 Percent Identity: 64.444

alignment_block:
US-09-668-482-2 x AL532444/rev ..

Align seg 1/1 to reverse of: AL532444 from: 1 to: 1028

222 ProIleAspValProPheSerGlyLeuTyrArgGlyLeuArgAlaArgAs 238
984 CCCATCGACTGCCCT...TCAGCGGCTGTACCGGGCATGTAGCGCGAAC 938
238 nPheIleHisSerLysIleGlnGluAsnIleArgLysLysIleGlnasp. 254
937 CTTAATTAAACGCGCGCATCGACAGAAAMATACGCCCAAGATCTGCGGGC 888
255 . AspAspAsnGluAsnGlnGln.LysTyrLysAspAlaLeuGlnLeuLe 270
887 TGCGGGCATCCGAGCGGCCCAAGGCTGCAGAAAGACGCCGTGACGTGTT 838
270 uIleGluAsnSerArgArgSerAspGluProPheSerLeuGlnAlaMetL 287
837 GATCGAGCAGCTCGTGGAGAGAGGGGAGAGCGGCTGACATGACAGGACTAA 788
287 ysgLualAlaThrGluLeuLeuPheGlyGlyHisGluThrThralaser 303
787 AGCAATCTTCAACCGAATCTCTTTGGAGGACACGAAACACAGGCCAGT 738
304 ThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGluValValGl 320
737 GCAGCCACATCTCTGATMACTTACCTGGGGCTCTTACCCACATGTTCTCCA 688
320 nLysValArgGluGluValGlnGluLysValGluMetGlyMetTyrThrP 337
687 GAAAGTGCAGAGAGAGCTGAAGAGTAAGGTTTACTTTGCAAGAGCAATC 638
337 roGlyLysGlyLeuSerMetGluLeuLeuAspGlnLeuLysTyrThrGly 353
637 AAGACAACAAGTTGACATGGAATTTTGGACAACCTTAATACATCGGG 588
354 CysValIleLysGluThrLeuArgIleAsnProProValProGlyGlyPh 370
587 TGTGTTATTAAAGGAGACCCCTTGACTGAATCCCCAGTTCACAGAGGTT 538
370 eArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysG 387
537 TCGGGTTCCTGAAGACTTTTGAATTAATGATACAGATTCCCAAGG 488
387 lYTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaAspVal 403
487 GCTGAATGTATCTACAGTATCTGTGATCTCATGATGTGGCAGAGATC 438
404 PheProAsnLysGluGluPheGlnProGluArgPheMetSerLysGlyLe 420
437 TTCACCAACAAGAAGAAATTATCTTCACCGATTTCATGCTGCCCTCACCC 388
420 uGluAspGlySerArgPheAsnTyrIleProPheGlyGlyGlySerArgM 437
387 AGAGGATGCATCCAGGTTTCAGCCTCATTTCCATTGGAGGAGGCCCTTAAGA 338
437 etCysValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeuVal 453
337 GCTGTGTAGCAAGAAATTGCAAAAATTTCTTCAAAAATATTACAGTG 288
454 GluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyProProThme 470
287 GAGCTGSCCAGGCAATGTSACTGCMGCTTCTTAATGACACCTTCAATAT 238
470 tLysThrGlyProThrIleTyrProValAspAsnLeuProThrLysPheT 487
237 GAAAMCYAGTCCACCGGTGATCTGTGACAAATCTCCCTGCAAGATTC 188
487 hrSerTyr 489
187 CCCATTTC 180


```
seq_name: gb_est29:AL539668

seq_documentation_block:
LOCUS      AL539668      974 bp      mRNA      EST      16-FEB-2001
DEFINITION AL539668 LTI_FL013_FBrn1 Homo sapiens cDNA clone CSODF036YG04 5
prime, mRNA sequence.
ACCESSION  AL539668
VERSION    AL539668
KEYWORDS  AL539668.1 GI:12869097
SOURCE    EST.
          human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 974)
AUTHORS   Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 Evry cedex - France
          Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
    source
        1. 974
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CSODF036YG04"
            /clone_lib="LTI_FL013_FBrn1"
            /dev_stage="pooled tissue from post conception fetuses (20
            week, 24 week and 26 week)"
            /lab_host="DH10B"
            /note="Organ: Fetal brain; Vector: PCMVSPORT 6; 1st strand
            cDNA was primed with a NotI-oligo(dt) primer. Five prime
            end enriched, double-stranded cDNA was digested with Not I
            and cloned into the Not I and Eco RV sites of the
            PCMVSPORT 6 vector. Library was constructed by Life
            Technologies. Contact : Feng Liang Life Technologies, a
            division of Invitrogen 9800 Medical Center Drive Rockville
            , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
            fliang@lifetech.com URL :
            http://fulllength.invitrogen.com"

BASE COUNT      223 a      269 c      305 g      177 t
ORIGIN
alignment_scores:
    Quality:      855.00      Length:      297
    Ratio:        3.476      Gaps:      5
    Percent Similarity: 82.828      Percent Identity: 63.636

alignment_block:
US-09-668-482-2 x AL539668 ..
Align seg 1/1 to: AL539668 from: 1 to: 974

61 ILEuGlnARgArgLysPheLeuARgMetLysARgGlnLysTYrGLyCY 77
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
95 CTGACCAAGCGGAGGAGAGTTCCTGCAGATGAAGCGCAGAAATACGGCTT 144
77 sILETyRlYsThRHisLeuPheGLyAsnProThrValARgValMetGLyA 94
   |||||
145 CATCTACAAGACGCACTGTTCGGCGGCCACCGCTACGGGTGATGGCG 194
94 laAspAsnValARgGlnILEuLeuGLyGLuHisLysLeuValSERVal 110
   |||||
195 CGGACAATGTGCGGCATCTTCTCGGAGAGAC.CGGCTGTGTGGGTC 243
111 GlnTrProlAsERValARgThrILEuGLySERAsprThrLeuSERAs 127
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
244 CACTGGCCAGCGTGGTGCACCACTTCTGGGATCTGGCTGCCCTCTCTAA 293
127 nValHisGLyValGlnHisLysAsnLysLysAlaILEuMetARgAlAP 144
   |:::|
```

```
294 CCTGCACGACTCCTCGCACAGCAGCGCAAGAAGTGTATATCGGGCCT 343
144 hESERARgAsPALALeUGlnHISTYrILEProVALILEGlnGLuVal 160
   |||||
344 TCAGCCGCGAGGA.CTCGAATGTACGTGCCGGGTGATCAGCAGGAAGTG 392
161 LysSERAlaILEGlnGLUTrPLEuGlnLysAspSERcys..... 173
   ||:::
393 GGCAGCAGCCTGAGCAGTGGCTG.....AGCTCGCGGAGCGCGG 433
174 .ValLeuValTYrProGLuMetLysLysLeuMetPheARgILEAlaMetA 190
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
434 CCTCGTGTCTACCCCGAGGTGAAGCGCCTCATGTTCGAAATCGCCATGC 483
190 rGILEuLeuGLyPheGLUProGLuGlnILELysThRAsp..... 203
   |||||
484 GCATCCTACTGGGCTGCGAACC...CAACTGGCGGGCGAGGGGACTCC 530
204 GLUGlnGLuLeuValGLuAlAPheGLuGLuMetILELysAsnLeuPheSE 220
   |||||
531 GAGCAGCAGCTTGTGAGAGCGCTTCGAGAAATGACCCGCAATCTTCTCTC 580
220 rLeuProLLeAspValAProphESERcLYLeuTYrARgLYLeuARgAlAA 237
   |||||
581 GCTGCCCATCGACGCTGCCCTTACGGGCTGTACCGGGCATGAAGCGC 630
237 rGAsnPheILEHisSERLysILEGLuGLuAsnILEARgLYsLILEGln 253
   |||||
631 GGAACCTCATTTCAACGCCGCGCATCGAGCAGAACAATTCCGCCCAAGATCTGC 680
254 Asp..AspAspAsnGLuAsnGLuGlnLysTYrLysAsPALALeUGlnLE 269
   ::::|
681 GGGCTGGCGGGCATCCGAGGGCGGCCAGGGCTGCAAAGACGCCCTGCACGCT 730
269 uLEuILEGLuAsnSERARgARgSERAspGLUProphESERLeuGlnAlAM 286
   |||||
731 GTTGATCGAGCACTCGTGGAGAGGGGAGAGCGGCTGACATCGACGGCAC 780
286 eTLysGLuAlaAlATHrGLuLeuLeuPheGLyGLYHISGLUTHrAlA 302
   ::|||
781 TAAAGCAATCTTCAACCGAAGCTCTTGGAGGACACAGAAACACCGGCC 830
303 sERThrAlATHrSERLeuValMetPheLeuGLyLeuAsnThrGLuValVA 319
   ||:::
831 AGTGCAGCCACACTCTGTGATCACTTACCTGGGGCTCTACCCACACATGTTCT 880
319 IGLnLysValARgGLuGLuValGLnGLuLysValGLuMetGLyMetTYRT 336
   ::|||
881 CCAGAAAGTGCAGAGAGAGCTGAAGAGTAAGGTTTACTTTGCAAGAGCA 930
336 hrProGLyLysGLyLeuSERMetGLuLeuLeuAspGlnLeu 349
   ::::
931 ATCAAGACAACAAGTTTGACATGGAATTTTGAAACACTTA 971

seq_name: gb_est71:BE189825

seq_documentation_block:
LOCUS      BE189825      696 bp      mRNA      EST      07-AUG-2000
DEFINITION db61c05.y1 wellcome CRC psk egg Xenopus laevis cDNA clone
IMAGE:3377480 5' similar to gb:gb|AF057566.1|AF057566 Xenopus
laevis retinotic acid converting enzyme (XENOPUS);, mRNA sequence.
ACCESSION  BE189825
VERSION    BE189825.2 GI:9729548
KEYWORDS  BE189825
SOURCE    EST.
          African clawed frog.
ORGANISM  Xenopus laevis
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
          Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 696)
AUTHORS   Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Page, D.,
          Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person
          , B., Gibbons, M., Harvey, N., Rlter, E., Jackson, Y., McCann, R.,
```


Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 119 a 145 c 136 g 129 t 1 others
ORIGIN

alignment_scores:
Quality: 792.00 Length: 154
Ratio: 5.176 Gaps: 0
Percent Similarity: 99.351 Percent Identity: 98.701

alignment_block:
US-09-668-482-2 x AI584636 ..

Align seg 1/1 to: AI584636 from: 1 to: 530

1 MetGlyLeuTyrThrLeuMetValThrPheLeuGlyThrIleValLeuPr 17
|||||
69 ATGGGGCTGTACACCCCTTATGTGCTACCTTCTCTGCACCATCGTCTACC 118
17 ovalleuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
|||||
119 CGTTTACTCTTTCTCGCCGGGTGAAGTGTGGAGATGTTAATGATCC 168
34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
|||||
169 GACGAGTCGATCCGAACACTGCAGAGTCTCTACCGCCAGGATACCATGGGC 218
5 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPhe 67
|||||
219 TTGCCGTTTCATTGGAGAAACGCTCCAGCTGATCTCCAGAGAGAAAGTT 268
67 eleuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeu 84
|||||
269 TCTGCCGATGAACGGCAGAAATACGGGTGCATCTACAAGCAGCACCCTCT 318
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|||||
319 TCGGGAACCCGACTGTGAGGTGATGGAGCTGATTAATGTGAGGCAGATT 368
101 leuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||
369 CTGCTGGGGCAGACACAGCTGTCTGTCTCAGTGGCCAGCATCAGTGAG 418
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||
419 AACCATCTCGGGCTGTGACACCCCTCTCAATGTCCATGGAGTTCAACACA 468
134 ySAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
|||||
469 AAAACAAGAAAAAGCCATTATNGAGGCGTCTCTCGAGATGCTCTGGAG 518
151 HisTyrIlePro 154
|||||
519 CACTACATTCCC 530

seq_name: gb_est43:AW171465

seq_documentation_block:
LOCUS AW171465 765 bp mRNA EST 12-NOV-1999
DEFINITION f142a06.x1 Sugano Kawakami zebrafish DNA Danio rerio CDNA clone
2640274 3' similar to SW:CP26_BRAKE P79739 CYTOCHROME P450 26 ;
ACCESSION AW171465
mRNA sequence.

VERSION AW171465.1 GI:6402990
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio

REFERENCE
AUTHORS Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S., Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T., Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
Washu Zebrafish EST Project 1999
Unpublished (1999)

TITLE JOURNAL
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: T7 ET from Amersham
High quality sequence stop: 514.
Location/Qualifiers

FEATURES
source

1.765
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="2640274"
/sex="mixed (one male and one female, including unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: PME18S-FL3; Site_1: DraIII (CACTGTG); Site_2: DraIII (CACCAGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested and cloned into distinct DraIII sites of the PME18S-FL3 vector (5' site CACTGTG, 3' site CACCAGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTCTGCTCTAAAGCTCG and 3' end primer CGACCTGCAGCTCGAGCACA."
BASE COUNT 225 a 174 c 154 g 212 t
ORIGIN

alignment_scores:
Quality: 765.50 Length: 171
Ratio: 4.845 Gaps: 1
Percent Similarity: 92.398 Percent Identity: 87.135

alignment_block:
US-09-668-482-2 x AW171465/rev ..

Align seg 1/1 to reverse of: AW171465 from: 1 to: 765

322 ValArgGluGluValGlnGluLysValGluMetGlyMetTyrThrProG1 338
|||||
765 GTCCACACAGCGGTTCAGACAGGTTAAAT...GGCCATGATACTCCGGT 719
338 yLysGlyLeuSerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCysV 355
|||||
718 CATAGGTTTGAGCATGAGCTGTGCACACAGCTGATACACTGGATGTG 669
355 aLleLysGluThrLeuArgIleAsnProProValProGlyGlyPheArg 371
|||||
668 TGAGTAAGAACCT.CTCAGAAATCAACCCCTCTGTCTCTGGAAGAATTCA 620

372 ValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTr 388
|||||:::||||| ||||||| ||||||| |||||
619 GTCGACATCGAACCCTCTGAATGATGGTTACCAATTCCTACAGGATG 570
388 pAsnValIleTyrSerIleCysAspThrHisAspValAlaAspValPheP 405
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
569 GAACGTCATTACAGCATCTGTGACACGACGATGTGGCCGACGCTTTC 520
405 roAsnLysGluGluPheGlnProGluArgPheMetSerLysGlyLeuGlu 421
|||||:::||||| ||||||| ||||||| ||||||| ||||||| |||||||
519 CAACAAGACGAGATTCAGCCGAGAGATTCATGAGCAAAAGTCTGAG 470
422 AspGlySerArgPheAsnTyrIleProPheGlyGlyGlySerArgMetCy 438
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
469 GACGGGTCCAGGTTTAACATACATCCCTTCGAGAGGATCCAGCATGTG 420
438 sValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeuValGluL 455
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
419 TGTGGCAAGAGTTCCGCAAAAGTGTACTCAAGATCTTTTAGTTGAGT 370
455 euThrGlnHisCysAsnTrpIleLeuSerAsnGlyProProThrMetLys 471
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
369 TAACGCAGCATGTCAATTGGATTCCTCAAAACGGACCCCGACAATGAAA 320
472 ThrGlyProThrIleTyrProValAspAsnLeuProThrLysPheThrSe 488
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
319 ACAGGCCCGACTGTTTACCAGTGACATCTCCCTACCAATTCACCTAG 270
488 rTyrrValArgAsn 492
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
269 TTATGTCAGAAAT 257

seq_name: gb_est48:AW513600

seq_documentation_block:

LOCUS AW513600 618 bp mRNA EST 03-MAR-2000
DEFINITION xo47b06.x1 NCI_CGAP_ut1 Homo sapiens cDNA clone IMAGE:2707091 3'
similar to SW:CP26_HUMAN O43174 CYTOCHROME P450 26 ;, mRNA
sequence.

ACCESSION AW513600 GI:7151678
VERSION AW513600.1
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 618)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
image.lnl.gov/image/html/iresources.shtml

FEATURES
source
Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco
High quality sequence stop: 440.
Location/Qualifiers
1..618

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2707091"
/clone_lib="NCI_CGAP_ut1"
/tissue_type="well-differentiated endometrial

adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"

BASE COUNT 111 a 191 c 199 g 115 t 2 others
ORIGIN
alignment_scores:
Quality: 739.50 Length: 207
Ratio: 4.131 Gaps: 4
Percent Similarity: 86.473 Percent Identity: 71.498

alignment_block:
US-09-668-482-2 x AW513600 ..

Align seg 1/1 to: AW513600 from: 1 to: 618

37 AspProAsnCysArgSerProLeuProProGlyThrMetGlyLeuProPh 53
|||:::||||| ||||||| ||||||| ||||||| ||||||| |||||||
10 GACCGAGTTGTGCCCTCCCAATGCCCCCGGAGCTATGGGCTTCCCTT 59
53 eIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPheLeuArgm 70
|:::||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
60 CTTTGGGGAACCTTGCAGATGTACTGCAGCGGAGAGAGTTCCTGCAGA 109
70 eTlyArgGlnLysTyrGlyCysIleTyrLysThrHisLeuPheGlyAsn 86
|||||:::||||| ||||||| ||||||| ||||||| ||||||| |||||||
110 TGAAGCCAGGAATACGGCTTCACTACAAGACCATCTGTCCGGCGG 159
87 ProThrValArgValMetGlyAlaAspAsnValArgGlnIleLeuLeuG 103
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
160 CCCACCGTACGGGTGATGGCGCGGACCAATGTGCCGGCCCATCTTGC 209
103 yGluHisLysLeuValSerValGlnTrpProAlaSerValArgThrIleL 120
|||||:::||||| ||||||| ||||||| ||||||| ||||||| |||||||
210 AGAGCACCGCGCTGTGTGCTGCTCACTGCGCAGCGTGGTGGCACCATTC 259
120 euGlySerAspThrLeuSerAsnValHisGlyValGlnHisLysAsnLys 136
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
260 TGGGATCTGGCTGCTCTTAACCTGCACGACTCCTCGCACAGCAGCGC 309
137 LysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlnHisTyrIl 153
|||||:::||||| ||||||| ||||||| ||||||| ||||||| |||||||
310 AAGAAGGTGATTATGCGGGCCTTCAGCCGCGAGGACACTCGAATGCTACGT 359
153 eProValIleGlnGlnGluValLysSerAlaIleGlnGluTrpLeuGlnL 170
:||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
360 GCCGGTGATCAACGAGGAAGTGGGCGACAGCCTGGAGCAGTGGCTG... 405
170 ysAspSerCys.....ValLeuValTyrProGluMetLysLys 182
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
406AGTCGCGCGGAGCGCGGCTCTGCTTACCCCGAGGTGAAGCGC 450
183 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGl 199
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
451 CTCATGTTCGGAATGCGCATGCGCATCTACTGTGTGCGAAGCC...CA 497
548 AATGACCCCGCAATCTTCTTCGCTGCCCATCGACGTGCNCTTCAGCGGG 597
199 nIleLysThrAsp.....GluGlnGluLeuValGluAlaPheGlu 213
|::: ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
498 ACTGGCGGGCGGCGGACTCCGAGCAGCAGCTTGTGAGGCGCTTCGAGG 547
213 LuMetIleLysAsnLeuPheSerLeuProIleAspValProPheSerGly 229
|||||:::||||| ||||||| ||||||| ||||||| ||||||| |||||||
548 AATGACCCCGCAATCTTCTTCGCTGCCCATCGACGTGCNCTTCAGCGGG 597
230 LeuTyrArgGlyLeuArgAla 236
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
598 CTGTACCGGGGCGATGAAGGCN 618
seq_name: gb_est51:AW765767


```
source
1. .847
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF036YG04"
/clone_lib="LTI_FL013_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Petal brain; Vector: PCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
PCMVSPORT 6 vector. Library was constructed by life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
filiang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      197 a      190 c      204 g      229 t      27 others
ORIGIN

alignment_scores:
Quality:      698.00      Length:      224
Ratio:      3.921      Gaps:      0
Percent Similarity: 79.464      Percent Identity: 60.714

alignment_block:
US-09-668-482-2 x ALS39667/rev ..

Align seg 1/1 to reverse of: ALS39667 from: 1 to: 847

264 lysaspalaleuGlnleuLeuIleGluasnSerArgSeraspGluPr 280
|||||
840 AAAGACGCGCTGCAGCTKTGTGATCGACACTCGTGGAGAGAGGAGAGCG 791
280 opheSerleuGlnAlaMetIlySGluAlaIatHrgluleuLeuPheGlyg 297
:::
790 GCTGGACATGACGGCATAAGCAATCTTCAACCGAATCTCTTTGGAG 741
297 lyHisgluThrThrAlaSerThrAlaThrSerleuValMetPheleugly 313
|||||
740 GACACGAACACGCGCCAGTGACGCCACATCTGTACCTTACCTGGGG 691
314 leuAsnThrGluValValGlnIlySValArgGluGluValGlnGluIlySva 330
|||||
690 CTCACCCACACACGCTCCCAAGAAAGCGGAGAGAGCTGAAGAGTAGGG 641
330 IgluMetGlyMetYrThrProGlyIlySerMetGluLeuA 347
:::
640 TTTACTTTGCAAGMCAATCAAGACAACAAGTTGACATGAAATTTTGG 591
347 spGlnleuIlySTyrThrGlyCysValIleIlySGluThrleuArgIleasn 363
:::
590 AACAACTCMATMCCNCGGTGTGTATTAAAGAGACCCCTCGACTGAAC 541
364 ProProValProGlyGlyPheArgValAlaLeuIlySThrPheGluLeuAs 380
|||||
540 CCCCCMGNCCAGGAGGGTTNCGGGNNGCTCCGAASMCTTCTGAATCAAA 491
380 ngIyTYrGlnIleProIlySGlyTPasnValIleTYrSerIleCysAspt 397
|||||
490 TGGATCCCMGANCCCCCAAGGCTGGAATGTATCTACAGTATCTGTGATA 441
397 hrHisaspValAlaaspValPheProasnIlySGluGluPheGlnProglu 413
|||||
440 CCCATGATGTGGCAGAGATCTTCAACCAAGAAGAAATTTAAACCTGAC 391
414 ArgPheMetSerIlySGlyLeuGluaspGlySerArgPheasnTYrIlePr 430
|||||
390 CGATTCATGTGCTGCCTCACCCMGAGGAGCAGACCCMGGCTCAGCCCATCC 341
430 opheGlyGlyGlySerArgMetCysValGlyIlySGluPheAlaIlySval 447
```

```
|||||
340 ATTTGAGAGAGGCGCTTAGAGCTGTGTAGGCAAGAAATTCGMAAAATTC 291
447 euleuIlyIlePheleuValGluLeuThrGlnHisCysasnTrpIleleu 463
|||||
290 CYCCAAATATATTACAGTGGAGCTGSCCAGGCAATGTGACTSGCCSCTR 241
464 SerasnGlyProProThrMetIlySThrGlyProThrIleTYrProvalAs 480
|||||
240 CTAATGAGACCTCTTACABGAAACACAGNCCACCGCTGTATCTGTGA 191
480 pasnLeuProThrIlySPheThr 487
|||||
190 CAATCTCCCTGCAGATTCACC 169

seq_name: gb_est101:BG515142

seq_documentation_block:
LOCUS      BG515142      571 bp      mRNA      EST      28-MAR-2001
DEFINITION dad60f10.y1 Wellcome CRC PCS107 tropicalis egg Silurana tropicalis
CDNA clone IMAGE:4464138 5' similar to TR:093323 093323 RETINOIC
ACID CONVERTING ENZYME. ;, mRNA sequence.
ACCESSION  BG515142
VERSION    BG515142.1 GI:13485799
KEYWORDS   EST.
SOURCE     western clawed frog.
ORGANISM   Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE  1 (bases 1 to 571)
AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
            Washu Xenopus EST project, 1999
            Unpublished (1999)
            Other ESTs: dad60f10.x1
            Contact: Sandy Clifton, Ph.D.
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by A. Zorn and J. Mason (Wellcome/CRC Institute
            ). DNA Sequencing by: Washington University Genome Sequencing
            Center
            Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 503.

FEATURES
source
1. .571
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="IMAGE:4464138"
/clone_lib="Wellcome CRC PCS107 tropicalis egg"
/tissue_type="egg"
/lab_host="DH10B (phage-resistant)"
/note="Vector: PCS107; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Average
insert size 1.5 kb, range 0.5-4 kb. Library constructed by
A. Zorn and J. Mason (Wellcome/CRC Institute). "
BASE COUNT      123 a      183 c      150 g      115 t
ORIGIN

alignment_scores:
Quality:      672.00      Length:      182
Ratio:      4.123      Gaps:      0
Percent Similarity: 89.560      Percent Identity: 67.033
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:25:24 ; Search time 89.59 Seconds
(without alignments)
336.311 Million cell updates/sec

Title: US-09-668-482-4
Perfect score: 2611
Sequence: 1 MGLPALIASALCTFVLPLL.....PTVYPVDNLPA RFTFHGEI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601: *
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT: *
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT: *
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT: *
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT: *
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT: *
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT: *
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT: *
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT: *
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT: *
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT: *
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT: *
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT: *
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT: *
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT: *
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT: *
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT: *
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT: *
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT: *
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT: *
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT: *
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT: *
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2611	100.0	497	19	AAW37734	Human cytochrome P
2	2611	100.0	497	19	AAW44160	Human retinoid met
3	2437	93.3	497	19	AAW37735	Cytochrome P450RAI
4	2437	93.3	497	19	AAW44161	Mouse retinoid met
5	1738	66.6	492	19	AAW37733	Cytochrome zP450RA
6	1738	66.6	492	19	AAW44159	Zebrafish retinoid
7	1113	42.6	216	21	AAB12489	Human PSEC64 prote
8	478	18.3	468	21	AAAG20783	Arabidopsis thalia
9	450.5	17.3	461	21	AAAG46490	Arabidopsis thalia
10	450.5	17.3	462	21	AAAG46489	Arabidopsis thalia
11	450.5	17.3	465	21	AAAG46491	Arabidopsis thalia

12	445.5	17.1	430	21	AAAG20784	Arabidopsis thalia
13	439	16.8	513	21	AAB07921	A cytochrome P450
14	434	16.6	481	21	AAAG30049	Arabidopsis thalia
15	434	16.6	489	21	AAAG30048	Arabidopsis thalia
16	433.5	16.6	472	18	AAW27153	Arabidopsis thalia
17	433.5	16.6	472	21	AAAG44571	Arabidopsis thalia
18	433.5	16.6	472	21	AAAG45022	Arabidopsis thalia
19	433.5	16.6	491	21	AAAG45021	Arabidopsis thalia
20	433.5	16.6	492	21	AAAG44570	Arabidopsis thalia
21	428	16.4	444	21	AAAG44572	Arabidopsis thalia
22	428	16.4	444	21	AAAG45023	Arabidopsis thalia
23	412.5	15.8	461	21	AAAG11836	Arabidopsis thalia
24	412.5	15.8	462	21	AAAG11835	Arabidopsis thalia
25	412.5	15.8	465	21	AAAG11834	Arabidopsis thalia
26	406.5	15.6	433	21	AAAG30050	Arabidopsis thalia
27	377	14.4	475	21	AAB07676	Amino acid sequenc
28	352	13.5	388	21	AAAG20785	Arabidopsis thalia
29	348.5	13.3	511	20	AAV23341	A P450-2 protein i
30	348.5	13.3	511	21	AAAB19694	Sweetgum coniferyl
31	335.5	12.8	513	22	AAAB31008	Amino acid sequenc
32	335.5	12.8	513	22	AAAB48181	B. napus F5H polyp
33	333	12.8	520	22	AAAB31007	Amino acid sequenc
34	333	12.8	520	22	AAB48179	B. napus F5H polyp
35	332	12.7	520	22	AAB48180	B. napus F5H polyp
36	327.5	12.5	163	21	AAAB40557	Human ORFX ORE321
37	322	12.3	520	18	AAW26640	Arabidopsis thalia
38	322	12.3	520	19	AAW40099	Arabidopsis ferula
39	322	12.3	520	21	AAAB15188	Arabidopsis ferula
40	320.5	12.3	520	21	AAAB23917	Arabidopsis thalia
41	313.5	12.0	488	21	AAAG13337	Arabidopsis thalia
42	313.5	12.0	488	21	AAAG1651	Arabidopsis thalia
43	313.5	12.0	527	21	AAAG13336	Arabidopsis thalia
44	310	11.9	471	21	AAAG23013	Arabidopsis thalia
45	310	11.9	479	21	AAAG23012	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AAW37734	
ID AAW37734	standard; Protein; 497 AA.
XX	
AC AAW37734;	
XX	
DT 07-JUL-1998	(first entry)
XX	
DE Human cytochrome P450RAI protein.	
XX	
KW Retinoid regulated gene; cytochrome P450 gene; enzyme;	
KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter.	
XX	
OS Homo sapiens.	
XX	
PN WO9749832-A2.	
XX	
PD 31-DEC-1997.	
XX	
PF 23-JUN-1997;	97WO-CA00488.
XX	
PR 01-OCT-1996;	96US-0724466.
PR 21-JUN-1996;	96US-0667546.
XX	
PA (TOOH) UNIV QUEENS KINGSTON.	
XX	
PI Petkovich PM;	
XX	
DR WPI: 1998-077193/07.	
DR N-PSDB; AAV09247.	
XX	
PT Identifying DNA encoding inducible or suppressible cytochrome P450 -	
PT by screening for drugs which reduce the catabolism of retinoic acid,	
PT useful in cancer chemotherapy and the treatment of acne and	

PT psoriasis
XX
PS Example 4; Pages 58-59; 113pp; English.
XX
CC This is the amino acid sequence of the human cytochrome P450RAI. Its
CC expression is dependent on the presence of retinoic acid (RA). The
CC retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RAI is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen
CC using promoters and nucleotide sequences to identify drugs which are
CC useful for reducing the catabolism of RA.
XX
SQ Sequence 497 AA;

Query Match 100.0%; Score 2611; DB 19; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.5e-253;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPDLLFLAIAIKLMDLYCVSGRDRSCALPLPPGTMGPFPFEGETLQM 60
Db 1 mglpallasalctfvlpdllflaiaiklwdlycvsgdrdrscalpippgtmgpfpfgetlqm 60
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGRPVTVRMGADNVRRILLGDDRLVSVHWPASVRTL 120
Db 61 vlqrrkflqmkrrkygfiythlfgrrptvrmgadnvrillgddrlvsvhwpasvrtll 120
QY 121 GSGCLSNLHDSHKKQKRVIMRAFSREALCEYVPVITTEVGSSLEQWLSGGERGLVYPE 180
Db 121 gsgclsnlhdsshkqrvimrafsrealceyvpviteevgssleqwlscgergllvype 180
QY 181 VKRLMFIAMRILIGCEPQLAGDGEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKAR 240
Db 181 vkrlmfiamrilligcepqlagdgsqqlveafeemtrnlfsldvpfsglyrgmkar 240
QY 241 NLIHARIEQNIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKQSTELLFG 300
Db 241 nlhariesnirakicglraseagggckdalqlliehswergerldmqalqkstellfg 300
QY 301 GHETTASAATSLITYLGLYPHVLQKVREELKSKGLCKSNQDNKLDMEILEQLKYIGCVI 360
Db 301 ghettasaatslitylglpvhvlokvreelkskglcksnqdnklhmeilleqlkyigcvi 360
QY 361 KETLRLNPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNPDRFS 420
Db 361 ketlrlnpvpggfrvalktfelngyqipkgnvlyisicdthdvaEIFtnkeefnpdrfs 420
QY 421 APHPEDASRFSFIPGGGLRSCVCGKEFAKILIKITVELARHCDWQLNGPPTMKTSPV 480
Db 421 aphpedasrfsfipggglrscvckefakilikitvelarhcdwqlngpptmktspv 480
QY 481 YPVDNLPARFTHHGEI 497
Db 481 ypvdnlparrfhfhei 497

RESULT 2
AAW44160
ID AAW44160 standard; Protein; 497 AA.
XX
AC AAW44160;
XX
DT 22-JUN-1998 (first entry)
XX
DE Human retinoid metabolising protein hp450RAI.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; human; inhibitor; antibody; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;

KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening.
XX
OS Homo sapiens.
XX
PN WO9749815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPT; 1998-077178/07.
DR N-PSDB; AAV12204.

PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 1; Page 54-55; 110pp; English.

This protein comprises a novel human retinoid metabolising protein,
designated hp450RAI. Its amino acid sequence was deduced from a
cDNA clone (see AAV12204) isolated from a retinoic acid-treated NT2
cell library. It includes a haem-binding motif characteristic of
cytochrome P450 proteins. hp450RAI is a retinoid oxidase that has
the ability to hydroxylate retinoic acid at the 4 position of the
beta-ionone ring, and is inducible in epithelial cells exposed to
retinoic acid. Zebrafish, human and mouse P450RAIs (see AAW44159-61)
are claimed. They can be expressed in host cells and used to
metabolize retinoic acid in an organism or cell, in drug screening,
and to raise antibodies useful for inhibiting retinoic acid
hydroxylation for the treatment of cancer, actinic keratosis, oral
leukoplakia, secondary tumours of the head and/or neck, non-small
cell lung carcinomas, basal cell carcinomas, acute promyelocytic
leukaemia, skin cancer, and premalignancy associated with actinic
keratosis, acne, psoriasis and/or ichthyosis.

SQ Sequence 497 AA;

Query Match 100.0%; Score 2611; DB 19; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.5e-253;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPDLLFLAIAIKLMDLYCVSGRDRSCALPLPPGTMGPFPFEGETLQM 60
Db 1 mglpallasalctfvlpdllflaiaiklwdlycvsgdrdrscalpippgtmgpfpfgetlqm 60
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGRPVTVRMGADNVRRILLGDDRLVSVHWPASVRTL 120
Db 61 vlqrrkflqmkrrkygfiythlfgrrptvrmgadnvrillgddrlvsvhwpasvrtll 120
QY 121 GSGCLSNLHDSHKKQKRVIMRAFSREALCEYVPVITTEVGSSLEQWLSGGERGLVYPE 180
Db 121 gsgclsnlhdsshkqrvimrafsrealceyvpviteevgssleqwlscgergllvype 180
QY 181 VKRLMFIAMRILIGCEPQLAGDGEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKAR 240
Db 181 vkrlmfiamrilligcepqlagdgsqqlveafeemtrnlfsldvpfsglyrgmkar 240
QY 241 NLIHARIEQNIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKQSTELLFG 300
Db 241 nlhariesnirakicglraseagggckdalqlliehswergerldmqalqkstellfg 300
QY 301 GHETTASAATSLITYLGLYPHVLQKVREELKSKGLCKSNQDNKLDMEILEQLKYIGCVI 360

Db 301 ghetasaatslityglyphvlqkvreelsksgllcksngdnkldmeileqlyigcvi 360
QY 361 KETLRNLNPVPGGFRVALKTFELNGYQIPKGMNVISICDTHDVAEIFTNKEEFNPDRES 420
Db 361 ketlrlnpvpvgfrvalktfelngyqipkgwnvysicdthdvaeiftnkeefnprfs 420
QY 421 APHPEDASRFSFIPEGGGLRSCVGKEFAKILKIFTVELARHCDWQLNGPPTMKTSPV 480
Db 421 aphpedasrfsfipfgglrscvgkefakillkiftvelarhcdwqlngpptmktstpv 480
QY 481 YPVDNLPARFTHFGEI 497
Db 481 ypvdnlparfthfgei 497

RESULT 3
AAW37735
ID AAW37735 standard; Protein; 497 AA.

AC AAW37735;
XX 07-JUL-1998 (first entry)
DE Cytochrome P450RAI isoform.
XX Retinoid regulated gene; cytochrome P450 gene; enzyme;
KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter; isoform.
OS Synthetic.
XX WO9749832-A2.

PN 31-DEC-1997.
PD 23-JUN-1997; 97WO-CA00488.
PF 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX (TOOH) UNIV QUEENS KINGSTON.
PA Petkovich PM;
XX WPI; 1998-077193/07.
DR N-PSDB; AAV09252.

XX Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis
PS Disclosure; Pages 59G-59H; 113pp; English.

XX This amino acid sequence is of an isoform of cytochrome zP450RAI.
CC Its expression is dependent on the presence of retinoic acid (RA).
CC The retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome P450
CC nucleotide sequence can be used to induce or suppress the expression
CC of its protein. P450RAI is highly induced by RA in cell lines and
CC tissues. This allows for development of a drug screen using promoters
CC and nucleotide sequences to identify drugs which are useful for
CC reducing the catabolism of RA.

XX Sequence 497 AA;

Query Match 93.3%; Score 2437; DB 19; Length 497;
Best Local Similarity 93.4%; Pred. No. 4.6e-236;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGLPALASALCTFVLLPLLLFLAAIKRLMDLYCVSGDRSCALPLPPTMGFPPEFGETLQM 60
Db 1 mglpallasalctfvlplllflaalklwdlycvssrdrscalplppgtmgfpffgetlqm 60
QY 61 VLQRRKFLQMKRRKRYGEIYKTHLFGRPYRWVGADNVRRIILGDDRLVSVHWPASVPTIL 120
Db 61 vlqrrkflqmkrrkrygyfkythlfgrrpyrwvgadnvrriilgddrlvsvhwpasvrtll 120
QY 121 GSGCLSNLHDSHKKRKKVIMRAFSREALECYVPVITEVGSLEQWLSCGERGLLVYPE 180
Db 121 gsgclsnlhdsstkrrkvvimrafsrealqcyvvlviaeevsscleqwlscgergllvype 180
QY 181 VKRLMFRIAMRILGCEPQLAGDGDSEQLVEAFEEMTNLFSLPIDVPFSGLYRGMKAR 240
Db 181 vkrImfriamrillgcepgpaggedeqlveafeemtrnlfsldvpfsglyrygvkar 240
QY 241 NLIHARIEONIRAKICGLRASEAGGCKDALQLLIEHSWGERGLDMQALKQSTELLFG 300
Db 241 nlharieeniraklrrlqatepdgycckdalqlliehswergerldmqalkgstellfg 300
QY 301 GHETTASAATSLITYGLYPHVLQKVREELSKGLCKSNQDNKLDMEILEQLKYGCVI 360
Db 301 ghetasaatslityglyphvlqkvreelskglcksngdnkldmetleqlyigcvi 360
QY 361 KETLRNLNPVPGGFRVALKTFELNGYQIPKGMNVISICDTHDVAEIFTNKEEFNPDRES 420
Db 361 ketlrlnpvpvgfrvalktfelngyqipkgwnvysicdthdvadiftnkeefnprf 420
QY 421 APHPEDASRFSFIPEGGGLRSCVGKEFAKILKIFTVELARHCDWQLNGPPTMKTSPV 480
Db 421 vphpedasrfsfipfgglrscvgkefakillkiftvelarhcdwqlngpptmktstpv 480
QY 481 YPVDNLPARFTHFGEI 497
Db 481 ypvdnlparfthfgei 497

RESULT 4
AAW44161
ID AAW44161 standard; Protein; 497 AA.

AC AAW44161;
XX 22-JUN-1998 (first entry)
DE Mouse retinoid metabolising protein mp450RAI.
XX Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antibody; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening.

OS Mus musculus.
XX WO9749815-A1.
PN 31-DEC-1997.
PD 23-JUN-1997; 97WO-CA00440.
PF 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX (TOOH) UNIV QUEENS KINGSTON.
PA Beckett BR, Jones G, Petkovich PM, White JA;

XX WPI; 1998-077178/07.
DR N-PSDB; AAV12205.
XX Retinoid metabolising protein - useful to develop products to treat,

PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT Ichthyosis
XX
PS Claim 1; Page 65-66; 110pp; English.
XX
CC This protein comprises a novel mouse retinoid metabolising protein,
CC designated mp450RAI. Its amino acid sequence was deduced from a
CC cDNA clone (see AAV12205) isolated from a retinoic acid-treated P19
CC teratocarcinoma library. It includes a haem-binding motif
CC characteristic of cytochrome P450 proteins. mp450RAI is a retinoid
CC oxidase that has the ability to hydroxylate retinoic acid at the 4
CC position of the beta-ionone ring, and is inducible in epithelial
CC cells exposed to retinoic acid. Zebrafish, human and mouse P450RAIs
CC (see AAW44159-61) are claimed. They can be expressed in host cells
CC and used to metabolize retinoic acid in an organism or cell, in drug
CC screening, and to raise antibodies useful for inhibiting retinoic
CC acid hydroxylation for the treatment of cancer, actinic keratosis,
CC oral leukoplakia, secondary tumours of the head and/or neck,
CC non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or Ichthyosis.
CC
XX
SQ Sequence 497 AA;

Query Match 93.3%; Score 2437; DB 19; Length 497;
Best Local Similarity 93.4%; Pred. No. 4.6e-236;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;
QY 1 MGIPALLASALCTFVLPPLFLAIAIKLMDLYCVSGRDRSCALPLPGTMGPFPGETLQM 60
Db 1 mglpallasalctfvlpplllflaalklwdlycvssrdscalplpgtmgpfpgetlqm 60
QY 61 VLQRRKFLQMKRRKYGFYKTHLFGPRTVRVMGADNVRRIILGDDRLVSVHWPASVRTIL 120
Db 61 vlqrrkflmqrrkygyfkythlfgprtvrmgadnvrriilgddrlvsvhwpasvrtil 120
QY 121 GSGCLSNLHDSHKKQKKVIMRAFSREALCYVPVITEEVGSSLEQWLSCGERGLVYPE 180
Db 121 gsgclsnlhds hkkqkkv imrafsrealcyv pvtteevgssleqwlscgerglvype 180
QY 181 VKRLMFRIAMRILGCEPOLAGDGDSEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKAR 240
Db 181 vkrlmfriamrilgcepolagdgds eqolveafeemtrnlfslpidvpfsglyrgmkar 240
QY 241 NLIHARIEQNIRAKICGLRASEAGQGCKDALQLLIEHSWGERLDMQALKOSSTELLEFG 300
Db 241 nlharieqnirakicglraseagqgckdalql l iehswgerldmqalkosstellf 300
QY 301 GHETTASATSLITVLYGLYPHVLOQVREELKSKGLCKSNQDNKLDMEILEQLKYIGCVI 360
Db 301 ghettasaatslityglyphvlqkvree l kskglcksnqdnkldmetleqlkyigcvi 360
QY 361 KETLRLNPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNDRFS 420
Db 361 ketlrlnpvpggfrvalktfelngyqipkgmnviysicdthdvaieftnkeefndrfs 420
QY 421 APHPEDASRFSFIPFGGLRSCVGEKFAKILKFTVELARHCDWQLNGPPTMKTSPV 480
Db 421 aphpedasrfsfipfgglrscvgekefak ilkftvelarhcdwqlngpptmktspv 480
QY 481 YPVDNLPAFTFHGEI 497
Db 481 ypvdnlpafthyfgdi 497

RESULT 5
AAW37733
ID AAW37733 standard; Protein; 492 AA.
XX
AC AAW37733;
XX
DF 07-JUL-1998 (first entry)

XX DE Cytochrome zP450RAI protein.
XX
KW Retinoid regulated gene; cytochrome P450 gene; enzyme;
XX oxidative metabolism; P450RAI; retinoic acid; RA; promoter.
OS Danio rerio.
XX
PN WO9749832-A2.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00488.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
PI Petkovich PM;
XX
DR WPI; 1998-077193/07.
DR N-PSDB; AAV09251.
XX
PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis
XX
XX Example 1; Pages 53-55; 113pp; English.

This is the amino acid for cytochrome zP450RAI of the zebra fish.
CC Its expression is dependent on the presence of retinoic acid (RA).
CC The retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RAI is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen using
CC promoters and nucleotide sequences to identify drugs which are useful
CC for reducing the catabolism of RA.
XX
SQ Sequence 492 AA;

Query Match 66.6%; Score 1738; DB 19; Length 492;
Best Local Similarity 68.0%; Pred. No. 7.5e-166;
Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;
QY 1 MGIPALLASALCTFVLPPLFLAIAIKLMDLYCVSGRDRSCALPLPGTMGPFPGETLQM 60
Db 1 mglytlmvtflctivlpvllflaavklwemlmirrvdpncrspilpgtmglpfigetlql 60
QY 61 VLQRRKFLQMKRRKYGFYKTHLFGPRTVRVMGADNVRRIILGDDRLVSVHWPASVRTIL 120
Db 61 vlqrrkflmqrrkygyfkythlfgprtvrmgadnvrriilgddrlvsvhwpasvrtil 120
QY 121 GSGCLSNLHDSHKKQKKVIMRAFSREALCYVPVITEEVGSSLEQWLSCGERGLV 177
Db 121 gsgclsnlhds hkkqkkv imrafsrealcyv pvtteevgssleqwlscgerglv 177
QY 178 YPEVKRLMFRIAMRILGCEP-OLAGDGDSEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRG 236
Db 177 ypevkr lmfriamrilgcep-olagdgds eqolveafeemtrnlfslpidvpfsglyrg 236
QY 237 MKARNLIHARIEQNIRAKICGLRASEAGQGCKDALQLLIEHSWGERLDMQALKOSSTE 296
Db 234 lrarnfihskieenlrkikgd-ddnenekykdaql lliensrdsdepfslqamkeate 292
QY 297 LIFGHETTASATSLITVLYGLYPHVLOQVREELKSKGLCKSNQDNKLDMEILEQLKYI 356
Db 293 llfghettastatslvmflglntevvgkvreeevgekvmgmytpgkglsmellldqlkyt 352

PS Claim 1; Page 10; 13pp; Japanese.

XX The present sequence represents a human protein, designated PSEC64, which
CC is related to neuron growth. The PSEC64 protein and its gene can be used
CC for the development of a preventive agent for use in the treatment of
CC diseases in which nerves are involved.

XX Sequence 216 AA;

Query Match 42.6%; Score 1113; DB 21; Length 216;
Best Local Similarity 99.1%; Pred. No. 1.4e-103;
Matches 214; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPALASALCTFVLPPLFLAIAIKLMDLYCVSGRDRSCALPLPGTMGFPEFTLQW 60
Db 1 mglpallasalctfvlplllflaiaiklwdlycvsgdrscalplpgtmgfifgetlqm 60
QY 61 VLQRRKFLQMKRRKYGFYKTHLFGRPVTRVMGADNVRILLGDDRLVSVHWPASVRTIL 120
Db 61 vlqrrkflqmkrrkygfylkthlfgprvtvmgadnvrillgghrlvsvhwpasvrtll 120
QY 121 GSGCLSNLHDSHKKQKKVIMAFSREALCYVPVITEVGSLSQWLSCGERGLVYPE 180
Db 121 gsgclsnlhdschkqkkvimsralsrealcyvpviteevgsleqwlscgergllype 180
QY 181 VKRLMFRIAMRILGCEPOLAGDGDSEQLVEAFEE 216
Db 181 vkrlmfriamrillgcepolagdgdsqqlveafee 216

RESULT 8

ID AAG20783 standard; Protein; 468 AA.

XX AAG20783;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 23102.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.

XX 06-SEP-2000.
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140623.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	17.38;	Score 450.5;	DB 21;	Length 461;
Best Local Similarity	27.8%;	Pred. No. 1.5e-36;		
Matches 136;	Conservative 96;	Mismatches 203;	Indels 55;	Gaps 15;
QY	1	MGIPALLASALCTFVLPILLFLAIIKWD--LYCVSGRDRSCALPLPGTGMFPFGETL	58	
Db	4	mglllllvs-lc-----sallrwnqmyrltng-----lpgltmgwplfgett	44	
QY	59	QMWLORRKFLQMKRRKYGFIYKTHLEGRPTVRVMGADNVRRIILGDDRLVSVHWPASVRT	118	
Db	45	eflkgqpnfmwrngrlrygsfkskhlgcptlismdsevnryilkneskglvpgypqsmld	104	
QY	119	ILSGCISNLHDDSSHQKKKIVIMRAFSREALCYV-PVITEEVGSSLEQWLSCGERGLL-	176	
Db	105	ilgtcmaavhygsshrlmrgslslslstmrhdilpkvdhfmrsyldqw---nelevid	161	
QY	177	VYPEVKRLMERIAMRILLGCEPOLAGDGDSEQOLVE---AFEEETRNLFSLPIDVPFSG	232	
Db	162	lqdktkhmaf-----lsslqtga--gnlrkpfveefktafkilvgvltisvpidipgtu	212	
QY	233	LYRGMKARNLTHARIEONIRAKIICGLRASEAGQGCKDALQLLIEHSWGERLDMOALKQ	292	
Db	213	yrqciqarnidrlrlrelmqer-----rdsgetftdmlylmkkegnrypltd-eeird	265	
QY	293	SSTELLFGGHETTASAAATSLITYLGLPVHVLQKVREELKSKGLCKSNODNKLDMEILEQ	352	
Db	266	qvvtlllysgyelvsttsmmalkylhdpkaiqlerae--hlafrerkrqdeplgledvks	323	
QY	353	LKIYGCIVIKETILRNPPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHVAEIFTNKE	412	
Db	324	mkftraviyetstrlativngvlrkttrdleiingyllipkgwriyyvtytreinydanlyedpl	383	
QY	413	EFNPDRFSAPHPEASRSEFIPIFGGGLRSCVKGKEFAKILLIKITVELARHCDQOLLNGPP	472	

Db 384 i f p m r w m k k s l e - s q n s c f v f g g t r l p c y k e l g i v e i s s f l n y f v t r y r w e e i g d e 441

QY 473 T M K T S P T V Y P 482

Db 442 l m ----- v f p 446

RESULT	10	
AAG46489		
ID	AAG46489	standard; Protein; 462 AA.
XX		
AC	AAG46489;	
XX		
DT	18-OCT-2000	(first entry)
XX		
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 58493.
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000;	2000EP-0301439.
XX		
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 17.3%; Score 450.5; DB 21; Length 462;
Best Local Similarity 27.8%; Pred. No. 1.5e-36;
Matches 136; Conservative 96; Mismatches 203; Indels 55; Gaps 15;

OY 1 MGLPALIASALCTFVLPLLFLAARIKD--LYCVSGRDRSCALPLPPGTMGFPEFGETL 58
Db 5 mgllllivs-lc-----salllrwqmrlytkng-----lppgtmgwplfgett 45

PR	22-JUL-1999;	99US-0145087
PR	22-JUL-1999;	99US-0145089
PR	22-JUL-1999;	99US-0145192
PR	23-JUL-1999;	99US-0145145
PR	23-JUL-1999;	99US-0145218
PR	23-JUL-1999;	99US-0145224
PR	26-JUL-1999;	99US-0145276
PR	27-JUL-1999;	99US-0145913
PR	27-JUL-1999;	99US-0145919
PR	28-JUL-1999;	99US-0145951
PR	02-AUG-1999;	99US-0146386
PR	02-AUG-1999;	99US-0146388
PR	02-AUG-1999;	99US-0146389
PR	03-AUG-1999;	99US-0147038
PR	04-AUG-1999;	99US-0147204
PR	05-AUG-1999;	99US-0147302
PR	05-AUG-1999;	99US-0147192
PR	05-AUG-1999;	99US-0147260
PR	06-AUG-1999;	99US-0147303
PR	06-AUG-1999;	99US-0147416
PR	09-AUG-1999;	99US-0147493
PR	09-AUG-1999;	99US-0147935
PR	10-AUG-1999;	99US-0148171
PR	11-AUG-1999;	99US-0148319
PR	12-AUG-1999;	99US-0148341
PR	13-AUG-1999;	99US-0148565
PR	13-AUG-1999;	99US-0148684
PR	16-AUG-1999;	99US-0149368
PR	17-AUG-1999;	99US-0149175
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	20-AUG-1999;	99US-0149723
PR	23-AUG-1999;	99US-0149929
PR	23-AUG-1999;	99US-0149930
PR	25-AUG-1999;	99US-0150566
PR	26-AUG-1999;	99US-0150884
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	30-AUG-1999;	99US-0151080
PR	31-AUG-1999;	99US-0151303
PR	01-SEP-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154039
PR	20-SEP-1999;	99US-0154779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156458
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	08-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0159293
PR	13-OCT-1999;	99US-0159294
PR	13-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159638
PR	18-OCT-1999;	99US-0159584
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768

PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	17.3%;	Score 450.5;	DB 21;	Length 465;
Best Local Similarity	27.8%;	Pred. No. 1.5e-36;		
Matches 136; Conservative	96;	Mismatches 203;	Indels 55;	Gaps 15;

```

QY      1 MGLPALLASALCTFVLPLLLFLAIAIKLWD--LYCVSGRDRSCALPLPPGTMGFPFFGETL 58
      |||  ::  |  ||  ::|:  |  |  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      8 mglIliivs-lc-----sallrwnqmqrytkng-----lppgtmgwpiifgeit 48

```

QY 59 QMVLORRKFLOMKRRKRYGFTLYKTHLEGRPTVRVMGADNVRILLGDRLVSVHWPASVRT 118
: : | : : : | : | : | : : | : : : : : :
Db 49 eflkgpnfmrngrlcygsffkshllgcptlismdsevnryillknesqglvpypgsmld 108

```

QY      119 ILGGCLSLNHDSSHKQKRVIMRAFSREALECYV-PVITEEVGSSLEQWLSCGERGLL- 176
      ||| :: :: ||| :: :: | :: :: :: :: :: :: :: :: :: :: :: ::
Db      109 llytcmavavhgsshrImrgslslsstmmrchlIpkdnfmrslYdqw--nelevId 165

```

```
QY      177 VYPEVKRLMERIAMRIILGCEPOLAGDGDSEQLVE-----AFEEEMTRNLFSLPIDVPPFG 232
        :   :   |   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      166 lqdktkhmaf-----lsslqtia--gnlrkpfveefktafklvgttlsvpidipgtn 216
```

```
QY 233 LYRGKARNLIHARIEONIRAKICGLRASEAGGCKDALQLLIEHSWGERGLDMQALQ 292
      |::||| | : : : |::| | |::| | : : :
Db 217 YRCGIQARNIDRLIRELMQER-----RDSGETFTCDMLGYLMKKEGNRYPLTD-EELRD 269
```

QY 293 STELLFGCHETASAAFTSLITYLGLYPHYLOKVRBELKSKGLLCKSNODNKLMEILEQ 352
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 270 qvvtilysgyevstsmmal kylndhpka lge lrae - hlafrerkrqdeplgledvks 327

```
QY 353 LKYIGCVIKETRLNPPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEITPNKE 412
      :|: || || || | | : |:||| |||| : : | :: :
Db 328 mkfraviyetsrlativngvlrkttrdleingylipkgrlyvytreinydanlyedpl 387
```

```

QY 413 EFNPDREFSAPHEDASRFSFIPIFGGGLRSCVCKEFAKILKIFTVELARHCDWOLINGPP 472
    ||| | : | : | ||| | ||| : : | : : |
Db 388 ifnpwrwmksle--sqnscfvggqtrlcpkgeliveissflhyfvttryrweigdde 445

```

QY	473	TMKTSPTVYP	482
	—	:	
Db	446	lm-----vIfp	450

RESULT	12
AA020784	
ID	AA020784 standard; Protein; 430 AA.

AC AAG20784;
XX
DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 23103.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;

XX Arabidopsis thaliana.
XX OS termination sequence.
KW gene expression control; promoter;
XX genetic mapping, gene expression control; promoter;
XX termination sequence.

XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

```

Query Match      17.1%; Score 445.5; DB 21; Length 430;
Best Local Similarity 27.3%; Pred. No. 4,4e-36;
Matches 124; Conservative 87; Mismatches 201; Indels 43; Gaps 11;

QY   49 MGPFPEGETLQWLOR-RKFLQMKRRKYGFITYKTHLEGRPTVRVMGADNVRRILLGGDRL 107
    |||:|||||: | |::|||:|||: | |: | :|:
Db   1 mgwpylgetllqysqnunvftskdkrygeifktrilglypcwmilaspeaarfvltvhnm 60

QY   108 VSVHWPASVRTLLGSGCLSNLDSSHQRKKVIMRAFSREALCYVPVITEEVSSTEGW 167
    :| | ::| | -| | -| ::| | :| | :| | :| |
Db   61 fkptypsrkeklippsalffhgdyshirkklvgssfyfetirklipdienhialsslgs 120

QY   168 LSCGERGLLYPEVKRMLFRIAMRILLGCEPOLAGDGDSQQLVAEFEEMTRNLF----- 222
    : | | :| | :| | :| | :| | :| | :| | :| |
Db   121 ammpi--vstygemkkfafdvglalafg-----hlessykelikhnyivdk 165

QY   223 ---SLPIDVPFSGLYRGMKARNLIHARIEQNIRAKICGLRASEAGQGCKDALQLIHSW 279
    | | :| | :| | :| | :| | :| | :| | :| |
Db   166 gynsfmslpjtsykhalkarmark---qlktivselicerrekralq--tdflghlnfkn 219

QY   280 ERGERLDMQALKOSSTELLFGGHETTASAATSLITTYGLYPHLQKVRELKS--KGLLC 337
    | :| | :| | :| | :| | :| | :| | :| | :| |
Db   220 ekgyvlteqjadnligvlfaagdttaclwlkyl---hddqklleavkaeqkalaYe 275

QY   338 KSNQDNK-IDMEILQULKYGVIKETLRLNPPVPGFRVALKTFFELNGYQIPKGMVIV 396
    ::::| | :| | :| | :| | :| | :| | :| | :| |
Db   276 ensrekkpltwrgtnmpltlhkviveslmasslstffreaavdvveykyllpkgwkmvp 335

QY   397 SICDTHDVAEIFTNKEEFNPDRFA-PHPEDASRFSEIFPGGLRSCVCGKEPAKLILKIT 455
    : | | :| | :| | :| | :| | :| | :| | :| |
Db   336 lfrnlhnprkyfsnpevfdsrfevnprkn-----tfmpfgsgvhacpnajaklgillit 390

```

```
QY 456 TVEARHCDWÖLLNGPPTMTSPTVYVVDNLPAF 490
      | : : : | : : : | : : : |
Db 391 lnhlsnfrwvkggkqigyspfipnglpatf 425
```

RESULT	13
AAB07921	
ID	AAB07921 standard; Protein; 513 AA
XX	
AC	AAB07921;
XX	
DT	14-NOV-2000 (first entry)
XX	

DE A cytochrome P450 enzyme designated DMF4.

KM DWF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
KW plant phenotype; cell elongation.
.....

05 Arabidopsis sp.

PN WO2000047715-A2.

PD 17-AUG-2000.

11-FEB-2000; 2000WO-US03820.

PR	11-FEB-1999;	99US-0119657.
PR	11-FEB-1999;	99US-0119658.

PA (ARIZ-) ARIZONA BOARD OF REGENTS.

PI Azpiroz R, Choe S, Feldmann KA;

DR WPI; 2000-549142/50.

PT New isolated dwf4 polynucleotide useful for altering the phenotype of
PT plants, for diagnostic assays and in the production of antibodies -

PS Claim 50; Fig 11; 113pp; English.

The present sequence represents a DMF4 polypeptide. The polypeptide is a cytochrome P450 enzyme that mediates multiple steps in synthesis of brassinosteroids. Specifically, it mediates multiple 2alpha-hydroxylation steps in brassinosteroid biosynthesis. The DMF4 polynucleotide is used for altering the phenotype of a plant. DMF4 plants display a dramatic reduction in the length of different organs, and this size reduction is attributable to a defect in cell elongation. The DMF4 polynucleotides and polypeptides can be used in diagnostic assays and to generate antibodies, which can be used to produce immunogenic compositions.

SQ Sequence 513 AA;

Query Match	16.88;	Score 439;	DB 21;	Length 513;
Best Local Similarity	27.08;	Pred. No. 2.5e-35;		
Matches 142; Conservative	85;	Mismatches 226;	Indels 72;	Gaps 15;

```
QY      3 LPALLASALCTFVLPPLLLFLAIAIKLMDLYCVSGRDRSCALPPLPGTMGFPFFEGETL---- 58
       ||||| : | ||||| : | ||||| : | ||||| : | ||||| : | ||||| : |
Db     10 lpllllpst-----lslllflillk-----rrnktlrlnlppgksqwpflgetiyglk 57
QY      59 -QMVLQRKRFLOMKRRKYGFITYKTHLEGRPTVRVMGADNVRRIILGGDRLVSVMHPASVR 117
       |:|   ||| ||:::||| ||:  | ||| : ||    : |||
Db     58 pytatltlgdfmqghvskkygkiyrsnlfgeptivsadaqlnrfilqnegr lfecysyrsig 117
QY      118 TILGSGLSNLHDDSSHKOKRKVIMRAFSEREALCYVPVITEEVGSS---LEOMUSCGER 173
        |||   : |   |:| : : : |   |   :: : :|   | : |
Db     118 gllygkwsmlylvqdmhrdmsislnfsharlr---tlilkdverhtl fvlidswqg--ns 172
QY      174 GLLVYPVKRLMFRIAMIRILLGCPEQLAGDGDSF-QOLVEAFEEMTRNLSLPIDVFPSG 232
```


Db 173 ifsaqdeakktftnlmakhimsmdp-----geeeteqlkkeyvtfmgkvnasaplnlpyta 227
QY 233 LYRGMKARNLIHARIEQNIRAKICGLRASEAGQGCKDALQLLIEHSWE-----RGERL 285
Db 228 yhkalgstratlIkfierkmeerklldikeed-----qeeevktedeaemsksdhvrkqrc 282
QY 286 DMQ---ALKQS---STE-----LLEFGHETTASATSLITFYGLYPHVLQKVREELK 331
Db 283 dddllgwlkhsnlstegldlilsllfaghetssvaialaiffiqacpkaveelreehl 342
QY 332 SKGLLCKSNQDNKIDMEILEQLKYICGYIKETLRLNPPVPGGFRVALKTFELNGYQIPKG 391
Db 343 eiarakkeligeseelnwddykkmdftgcvinetlrlgnvvrflhkalkdvrrykgydipsg 402
QY 392 WNVYISICDTHDYAEIFTNKEEFNPDRFSAPHEDASRFS-----FIFPGGLRSC 442
Db 403 wkvlpvisavhldnsrydqpnlfnprwqgqngnasssgsfstcwnqnympfgg99prlc 462
QY 443 VGKEFAKILKIFTVELARHCDWQLLNGPPTMKTSPTYPVVDNLP 487
Db 463 agselaklemavflhhlvlkfnwelae-----ddqpfaifpfvdfp 502

RESULT 14
AAG30049
ID AAG30049 standard; Protein; 481 AA.
XX AAG30049;
AC AAG30049;
XX 17-OCT-2000 (first entry)
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 35855.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35855.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX EP1033405-A2.
PN EP1033405-A2.
XX 06-SEP-2000.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 06-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.6%; Score 434; DB 21; Length 481;
Best Local Similarity 28.0%; Pred. No. 7.4e-35;
Matches 140; Conservative 80; Mismatches 248; Indels 32; Gaps 14;

QY 1 MGLPALASALCTFVLPDLLFLAIAIKLMDLYCVSGRDRSCALPLPPTMGFPFEGTLQM 60
Db 1 mwfp111---lg1fv1kwv--lkrvnw1yvs1gk1gek1hy---lppd1gwpv1gnmwsf 52
QY 61 V-----LQRRKFLQMKRRKYG--FIYKTHLFGRTVRVMGADNVRRI1LGDRLVSVHWP 113
Db 53 lra1ktsdps1f1qsy1ltrygrtgy1kahmfgypcvlvt1petcrrv1tddaf-h1gwp 111
QY 114 ASVRTILSGSGLNLHDSHKQRRKVIMRAFS-REALECYVPVITTEYGSLSLEQWLSCE 172
Db 112 kstmk1igrkstfv1gisfeehkrl1rltsapvngpealsvy1qf1eetvnt1lekws1kme 171
QY 173 RGLLVPEVVKRLMFRIAMRI1LGCPEQLAGDGDSEQQLVEAFEEETRN1LFSLPIDVPFSG 232
Db 172 ief1sh--lrlkt1fkvimy1flssesehvm1d----slerey1nl1ygvramg1n1pgfa 224
QY 233 LYRGMKARNL1HARIEQONIRAKICGLRASEAGQCKDALQ1LLIEHSWERGER1DMQALKQ 292
Db 225 yhralkar1k1vaaf-gsivtnrrn1qrknissnrk1md1dl1dvk1dengrv1ddeel1d 283
QY 293 SSTE1LFGGHETTASAATSL1TY1GLYPHV1QKVREELKSKGLCKSNQDNK1DME1LEQ 352
Db 284 111my1nagnessgh1tmwat1lmqehpm1lqkakee--ger1vkkrapgqk1ltketre 341
QY 353 LKYIGCV1KETRLN1PVPVPGGF1RYALKTFE1LNGYQ1PKGWN1Y1SICDTHDVAE1FTNKE 412
Db 342 mvyl1sgv1det1rlv1t1fs1ta1freaksdv1qmdgy1l1pk1gwk1v1tw1frnv1ld1pel1ydp1k 401
QY 413 EFNPDR1SAPH1EDASRF1FPGG1RSCV1GKEF1AK1L1K1FTV1ELARHCDWQ1LN-GP 471
Db 402 k1dps1wegy1t1pkag---t1flp1f1g1sh1c1p1gnd1ak1le1st1f1nh1fl1k1yrv1ersn1pgc 458
QY 472 PTMK1TSPTVY1PVDN1LPA1RFT 491
Db 459 pvm-f1lphnr1pkdncl1arit 477

RESULT 15
AAG30048
ID AAG30048 standard; Protein; 489 AA.
AC AAG30048;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35854.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.6%; Score 434; DB 21; Length 489;
Best Local Similarity 28.0%; Pred. No. 7.5e-35;
Matches 140; Conservative 80; Mismatches 248; Indels 32; Gaps 14;

OY 1 MGLPALLASALCTFVLPPLFLAIIKIWDLYCVSGRDRSCALPPLPGTMGPFPGETLQM 60
Db 9 mwfpflll---lglfvlkvw--lkrvnwlyvsklgekkhy---lppgdlgwvpiqnmwsf 60
OY 61 V-----LQRRKFLQMRRKYG--FIYKTHLFGRPTRVWGADNVRRILLGDDRIVSVHWP 113
Db 61 lrafktsdpesfiqsyitrygrgiykahmfgypcvltvpetcrrvltddaf-higwp 119
OY 114 ASVRTILGSGCLSNLHDSHROKKVIMRAFS-REALECYVPVITEEVGSSLEQWLSCGE 172
Db 120 kstmkligrksfvgisfeehkrlrltsapvngpealsvylqfieetvntdlekwskmge 179
OY 173 RGLLVYPEVKRLMFRIAMRILGCEPQLAGDGDSEQQLVEAFEEEMTRNLFSLPIDVPPSG 232
Db 180 iefish--lrkltfkvimylfissesehvmd-----slereytnlnygvramglnlpgfa 232
OY 233 LYRGMKARNLIHARIEQNTIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKQ 292
Db 233 yhralkarkklvaaf-qslvtnrrnrgkrqnissnrkmdlnlidvkdengrvlddeeiid 291
OY 293 STELLFGHETTASATSLITYLGLYPHYLVQKVRRELKSKGLLCKSNQDNKLDMEILEQ 352
Db 292 llmlylnagnessghltnwatilmqehpmilqakee--qerivkkrapgqkltlketre 349
OY 353 LKYIGCVIKETLRLNPNYPGGERVALKTFELNGYQIPKGNVYISICDTHDVAEIFTNKE 412
Db 350 mvyisqvidetlrvltfsltafreaksdvqmdgyilpkrqkwvltwfrnvhlrdpeiydpk 409
OY 413 EFNPDRESAPHPEDASRFSFIPFGGLRSCVCGKEFAKILKIFVELARHCDWQLN-GP 471
Db 410 kfdpstrwegytpkag---tflpfglgshlcpndlakleisiflhflkkyrversnpgc 466
OY 472 PTMKTSPTVYPVDNLPARFT 491
Db 467 pvm-flphnrpkdnclarit 485

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:27:04 ; Search time 88.14 Seconds
(without alignments)
126.891 Million cell updates/sec

Title: US-09-668-482-4
Perfect score: 2611
Sequence: 1 MGLPALASALCTFVLPPLL.....PTVYPVDNLPARTFHGEI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2611	100.0	497	3	US-08-724-466B-4 Sequence 4, Appli
2	2611	100.0	497	4	US-08-882-164D-4 Sequence 4, Appli
3	2437	93.3	497	4	US-08-882-164D-32 Sequence 32, Appli
4	1738	66.6	492	3	US-08-724-466B-2 Sequence 2, Appli
5	1738	66.6	492	4	US-08-882-164D-2 Sequence 2, Appli
6	433.5	16.6	472	2	US-08-622-166A-2 Sequence 2, Appli
7	433.5	16.6	472	2	US-08-622-166A-4 Sequence 4, Appli
8	348.5	13.3	511	4	US-08-991-677-4 Sequence 4, Appli
9	322	12.3	520	2	US-09-091-432-2 Sequence 2, Appli
10	289	11.1	576	3	US-08-948-564-16 Sequence 16, Appli
11	288.5	11.0	510	3	US-08-948-564-4 Sequence 4, Appli
12	271.5	10.4	513	3	US-08-948-564-2 Sequence 2, Appli
13	264.5	10.1	496	1	US-08-313-075A-50 Sequence 50, Appli
14	252.5	9.7	500	4	US-09-292-768-68 Sequence 68, Appli
15	252.5	9.7	500	4	US-09-292-768-70 Sequence 70, Appli
16	251.5	9.6	500	4	US-09-292-768-4 Sequence 4, Appli
17	251	9.6	510	3	US-08-606-505B-66 Sequence 66, Appli
18	251	9.6	510	4	US-09-616-990-66 Sequence 66, Appli
19	249	9.5	504	1	US-08-457-274A-25 Sequence 25, Appli
20	249	9.5	504	5	PCT-US95-05758-25 Sequence 25, Appli
21	246.5	9.4	500	3	US-08-881-784-9 Sequence 9, Appli
22	242.5	9.3	517	1	US-08-457-274A-2 Sequence 2, Appli
23	242.5	9.3	517	5	PCT-US95-05758-2 Sequence 2, Appli
24	240	9.2	496	4	US-09-292-768-64 Sequence 64, Appli
25	239.5	9.2	517	1	US-08-457-274A-28 Sequence 28, Appli
26	239.5	9.2	517	5	PCT-US95-05758-28 Sequence 28, Appli
27	238.5	9.1	476	1	US-08-313-075A-30 Sequence 30, Appli

28	238.5	9.1	532	3	US-08-948-564-10	Sequence 10, Appli
29	237	9.1	498	1	US-08-457-274A-24	Sequence 24, Appli
30	237	9.1	498	5	PCT-US95-05758-24	Sequence 24, Appli
31	237	9.1	509	3	US-08-948-564-18	Sequence 18, Appli
32	235	9.0	496	3	US-08-881-784-1	Sequence 1, Appli
33	235	9.0	496	4	US-09-292-768-2	Sequence 2, Appli
34	235	9.0	500	2	US-08-314-601-2	Sequence 2, Appli
35	235	9.0	500	5	PCT-US95-13051-2	Sequence 2, Appli
36	234.5	9.0	513	3	US-08-948-564-6	Sequence 6, Appli
37	234	9.0	496	4	US-09-292-768-66	Sequence 66, Appli
38	232	8.9	523	3	US-08-606-505B-67	Sequence 67, Appli
39	232	8.9	523	4	US-09-616-990-67	Sequence 2, Appli
40	231.5	8.9	508	4	US-08-991-677-2	Sequence 2, Appli
41	228	8.7	526	1	US-08-298-426-4	Sequence 4, Appli
42	227.5	8.7	496	4	US-09-172-339-6	Sequence 6, Appli
43	225	8.6	507	1	US-08-457-274A-23	Sequence 23, Appli
44	225	8.6	507	5	PCT-US95-05758-23	Sequence 23, Appli
45	224.5	8.6	490	1	US-08-201-118-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-724-466B-4
; Sequence 4, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-724-466B-4

Query Match 100.0%; Score 2611; DB 3; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.3e-279;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPALASALCTFVLPPLLFLAIAIKLMDLYCVSGRDRSCALPLPPTMGFFFGFTLQM 60
Db 1 MGLPALASALCTFVLPPLLFLAIAIKLMDLYCVSGRDRSCALPLPPTMGFFFGFTLQM 60

```

QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGPRTVRVMGADNVRILLGDDRLVSVHWPASVRTIL 120
|
|
|
Db 61 VLQRRKFLQMKRRKYGFIYKTHLFGPRTVRVMGADNVRILLGDDRLVSVHWPASVRTIL 120
QY 121 GSGCLSNLHDSHKKQKKVIMRAFSREALCYVPVITEEVGSSLEQWLSGGERGLVYPE 180
|
|
|
Db 121 GSGCLSNLHDSHKKQKKVIMRAFSREALCYVPVITEEVGSSLEQWLSGGERGLVYPE 180
QY 181 VKRLMFIAMRILGCEPQLAGDGDSEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGMKAR 240
|
|
|
Db 181 VKRLMFIAMRILGCEPQLAGDGDSEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGMKAR 240
QY 241 NLIHARIEQONIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKQOSTELLFG 300
|
|
|
Db 241 NLIHARIEQONIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKQOSTELLFG 300
QY 301 GHETTASATSLITYLGLYPHYLVQKVREELKSKGLCKSNQDNKLDMEILEQLKYIGCVI 360
|
|
|
Db 301 GHETTASATSLITYLGLYPHYLVQKVREELKSKGLCKSNQDNKLDMEILEQLKYIGCVI 360
QY 361 KETLRINPVPVPGGFRAVAKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNDRFS 420
|
|
|
Db 361 KETLRINPVPVPGGFRAVAKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNDRFS 420
QY 421 APHPEDASRFSFIPEGGLRSCVCGKEFAKILKIFTVELARHCDWQLNGPPTMTSPTV 480
|
|
|
Db 421 APHPEDASRFSFIPEGGLRSCVCGKEFAKILKIFTVELARHCDWQLNGPPTMTSPTV 480
QY 481 YPVDNLPARFTFHGEI 497
|
|
|
Db 481 YPVDNLPARFTFHGEI 497

```

```

RESULT 2
US-08-882-164D-4
; Sequence 4, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-882-164D-4

```

```

Query Match 100.0%; Score 2611; DB 4; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.3e-279;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MGLPALLASALCTFVLPILLFLAIAIKLMDLYCVSGDRSCALPLPPTMTGFPFFGETIOM 60
|
|
|
Db 1 MGLPALLASALCTFVLPILLFLAIAIKLMDLYCVSGDRSCALPLPPTMTGFPFFGETIOM 60
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGPRTVRVMGADNVRILLGDDRLVSVHWPASVRTIL 120
|
|
|
Db 61 VLQRRKFLQMKRRKYGFIYKTHLFGPRTVRVMGADNVRILLGDDRLVSVHWPASVRTIL 120
QY 121 GSGCLSNLHDSHKKQKKVIMRAFSREALCYVPVITEEVGSSLEQWLSGGERGLVYPE 180
|
|
|
Db 121 GSGCLSNLHDSHKKQKKVIMRAFSREALCYVPVITEEVGSSLEQWLSGGERGLVYPE 180
QY 181 VKRLMFIAMRILGCEPQLAGDGDSEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGMKAR 240
|
|
|
Db 181 VKRLMFIAMRILGCEPQLAGDGDSEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGMKAR 240
QY 241 NLIHARIEQONIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKQOSTELLFG 300
|
|
|
Db 241 NLIHARIEQONIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKQOSTELLFG 300
QY 301 GHETTASATSLITYLGLYPHYLVQKVREELKSKGLCKSNQDNKLDMEILEQLKYIGCVI 360
|
|
|
Db 301 GHETTASATSLITYLGLYPHYLVQKVREELKSKGLCKSNQDNKLDMEILEQLKYIGCVI 360
QY 361 KETLRINPVPVPGGFRAVAKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNDRFS 420
|
|
|
Db 361 KETLRINPVPVPGGFRAVAKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNDRFS 420
QY 421 APHPEDASRFSFIPEGGLRSCVCGKEFAKILKIFTVELARHCDWQLNGPPTMTSPTV 480
|
|
|
Db 421 APHPEDASRFSFIPEGGLRSCVCGKEFAKILKIFTVELARHCDWQLNGPPTMTSPTV 480
QY 481 YPVDNLPARFTFHGEI 497
|
|
|
Db 481 YPVDNLPARFTFHGEI 497

```

```

RESULT 3
US-08-882-164D-32
; Sequence 32, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996

```

APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-32

Query Match 93.3%; Score 2437; DB 4; Length 497;
Best Local Similarity 93.4%; Pred. No. 3.7e-260;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGLPALASALCTFVLPFLFLAATKIMDLVYSGRDRSCALPLPPTMGFPFEGETLQ 60
DB 1 MGLPALASALCTFVLPFLFLAATKIMDLVYSGRDRSCALPLPPTMGFPFEGETLQ 60
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGPTVRVMGADNVRRIILGDDRLVSVHWPASVRTIL 120
DB 61 VLQRRKFLQMKRRKYGFIYKTHLFGPTVRVMGADNVRRIILGDDRLVSVHWPASVRTIL 120
QY 121 GSGCLSNLHDSHKQKKVIMRAFSREALCYVPVITEEVGSSLEQMLSCGERGLVYPE 180
DB 121 GAGCLSNLHDSHKQKKVIMQAFSREALQCYVLVIAEVSLSCLQMLSCGERGLVYPE 180
QY 181 VKRLMFRIAMRILGCEPOLAGDSEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKAR 240
DB 181 VKRLMFRIAMRILGCEPOLAGDSEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKAR 240
QY 241 NLIHARIEONIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKOSSTELLFG 300
DB 241 NLIHARIEONIRAKIRLQATEPDGCKDALQLLIEHSWGERLDMQALKOSSTELLFG 300
QY 301 GHETTASAATSLITYLGLYPHVLOKVREELSKGLLCKSNQDNKLDMEILLEQLKYICVI 360
DB 301 GHETTASAATSLITYLGLYPHVLOKVREELSKGLLCKSNQDNKLDMEILLEQLKYICVI 360
QY 361 KETLRINPVPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEFTNKKEFNPDREFS 420
DB 361 KETLRINPVPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEFTNKKEFNPDREFS 420
QY 421 APHPEDASRFSFIPEGGLRSCVCKEFAKILKIFTVELARHCDWQLNGPPTMTKTSPTV 480
DB 421 VPHPEDASRFSFIPEGGLRSCVCKEFAKILKIFTVELARHCDWQLNGPPTMTKTSPTV 480
QY 481 YPVDNLPARFTFHGEI 497
DB 481 YPVDNLPARFTYFOGDI 497

RESULT 4
US-08-724-466B-2
Sequence 2, Application US/08724466B
Patent No. 6063606
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
ZIP: M5L 1A9

COUNTRY: Canada
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-2

Query Match 66.6%; Score 1738; DB 3; Length 492;
Best Local Similarity 68.0%; Pred. No. 5.3e-183;
Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;

QY 1 MGLPALASALCTFVLPFLFLAATKIMDLVYSGRDRSCALPLPPTMGFPFEGETLQ 60
DB 1 MGLYTLMTVFLCTIVLPVLLFLAAVKLWEMLMIRVDPNCRSPPLPPTMGFLPIGETLQL 60
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGPTVRVMGADNVRRIILGDDRLVSVHWPASVRTIL 120
DB 61 ILQRRKFLMKRQKYGCIYKTHLFGNPTVRVMGADNVRQIILGEHKLVSQVMPASVRTIL 120
QY 121 GSGCLSNLHDSHKQKKVIMRAFSREALCYVPVITEEVGSSLEQWL--SCGERGLV 177
DB 121 GSDTLSNVHGVOHKKKKAIMRAFSRDALHXYIPVIOQEVKSAIQEWLQKDSCL--VLV 176
QY 178 YPEVKRLMFIAMRILGCEP-OLAGDSEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRG 236
DB 178 YPEVKRLMFIAMRILGFEPEQLKTD--EQELVEAFEEEMIKNLFSLPIDVPFSGLYRG 233
QY 237 MKARNLIHARIEONIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKOSSTE 296
DB 234 LRARNFHSKIEENIRKKIQD-DNENEQKKDALQLLIEHSRSDSEFSLQAMKEATE 292
QY 297 LFFGHETTASAATSLITYLGLYPHVLOKVREELSKGLLCKSNQDNKLDMEILLEQLKYI 356
DB 293 LFFGHETTASTATSLVMFLGLNTEVVQKVREEVQEKVEGMNTPGKGLSMELLQKLYT 352
QY 357 GCYIKETLRINPVPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEFTNKKEFN 416
DB 353 GCYIKETLRINPVPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEFTNKKEFN 412
QY 417 DRFSAPHPEDASRFSFIPEGGLRSCVCKEFAKILKIFTVELARHCDWQLNGPPTMTK 476
DB 413 ERFMSKGLSDGSRFNYIPFGGSGRMCVCKEFAKVLKIFVELTQHCNWLNSGPPTMTK 472
QY 477 SPTVYPVNDLPARFTHF 493
DB 473 GPTIYPVNDLPTKFTSY 489

RESULT 5
US-08-882-164D-2
Sequence 2, Application US/08882164D
Patent No. 6306624


```

QY      252 RAKICGLR--ASEAGOGCKDALÖLLI--EHSWGERGLDMQALKÖSSTELLEFGCHETTAS 307
      :      |      |      |      |      |      |      |      |      |      |
Db      232 TVVVMKRREEEEEGAERKKMDLALALLAADGFSDEEIVDF-----LVALLVAGYETTST 285

QY      308 AATSLITVGLYPHVLÖKVREELKSKGLLCKSNÖDNKLDMEITLÖLKYGVCVIKETLRIN 367
      |      |      |      |      |      |      |      |      |      |
Db      286 IMTLAVKFLETTETPLALAÖLKEE--HEKIRAMKSDSYLEWSDYKSMPTÖCVVNETLRVA 343

QY      368 PVPGFGRVALKTEFLNGYÖIPKGMNVIYSICDTHDVAEIFTNKEEFNPDREFSAPHBEDA 427
      :      |      |      |      |      |      |      |      |      |
Db      344 NIIGGVFRAMTDVEIKGYKIPKGMKVYSSFRAVHLDPNHFKARTENPWRWÖSNSVYTG 403

QY      428 SRFSFIPEGGGLRSCVKGKEFAKILKLETVELARHCDWÖLNLNGPPTMKTSPTVYVPVDNLP 487
      |      |      |      |      |      |      |      |      |      |
Db      404 PSNVFTPEGGGRCLCPGYELARVALSVFLHRLVTGFSW-----VPAÖDKLVFFPTTRQ 458

QY      488 ARETHF 493
      |      |
Db      459 KRYPIF 464

```

```

1      RESULT 7
2      US-08-622-166A-4
3      ; Sequence 4, Application US/08622166A
4      ; Patent No. 5952545
5      ;
6      ; GENERAL INFORMATION:
7      ;
8      ; APPLICANT: KONCZ, CSABA
9      ; APPLICANT: MATHUR, JAIDEEP
10     ; APPLICANT: SZEKERES, MIKLOS
11     ; APPLICANT: ALTMANN, THOMAS
12     ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
13     ; TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
14     ; NUMBER OF SEQUENCES: 4
15     ; CORRESPONDENCE ADDRESS:
16     ;
17     ; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
18     ; STREET: P.O. Box 747
19     ; CITY: Falls Church
20     ; STATE: Virginia
21     ; COUNTRY: USA
22     ; ZIP: 22040-0747
23     ;
24     ; COMPUTER READABLE FORM:
25     ; MEDIUM TYPE: Floppy disk
26     ; COMPUTER: IBM PC compatible
27     ; OPERATING SYSTEM: PC-DOS/MS-DOS
28     ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
29     ;
30     ; CURRENT APPLICATION DATA:
31     ;
32     ; APPLICATION NUMBER: US/08/622,166A
33     ; FILING DATE: 27-MAR-1996
34     ;
35     ; CLASSIFICATION: 800
36     ;
37     ; ATTORNEY/AGENT INFORMATION:
38     ;
39     ; NAME: SVENSSON, LEONARD R.
40     ; REGISTRATION NUMBER: 30,330
41     ; REFERENCE/DOCKET NUMBER: 0147-0153P
42     ; TELECOMMUNICATION INFORMATION:
43     ;
44     ; TELEPHONE: (703) 205-8000
45     ; TELEFAX: (703) 205-8050
46     ;
47     ; TELEX: 248345
48     ;
49     ; INFORMATION FOR SEQ ID NO: 4:
50     ;
51     ; SEQUENCE CHARACTERISTICS:
52     ;
53     ; LENGTH: 472 amino acids
54     ; TYPE: amino acid
55     ; TOPOLOGY: linear
56     ; MOLECULE TYPE: protein
57     ;
58     ; US-08-622-166A-4

```

Query Match	16.6%;	Score 433.5;	DB 2;	Length 472;
Best Local Similarity	26.5%;	Pred. No. 5e-39;		
Matches 129; Conservative	90;	Mismatches 228;	Indels 39;	Gaps 133;

```
QY      19 LLELAIKLMDLYCVSGDRDSCALPLPGTIMGFPEGETLOWV-----LÖRRKETÖMKRR 73
```

Db	7	LLLLSSIAAGFLLLLL - RTRYRRMGDPGSLGPLIGETTÖLLIGA YKTE NPPEPFI DERVA	65
QY	74	KYGFYKTHLEGRPTVRVMGADNVRILLGGDRLVSVHMPASVRTILGSGCLSNHDSH	133
Db	66	RYGSVFMTHEGEPTIFSADPETNREVLÖNEGKLFECSPASICNLLGKHSLLMKSLH	125
QY	134	KÖRKKVIMRAFSREALLECYPV - ITEEVGSLEÖWLSCGERGLLVPEVKRLMFRIAMRI	192
Db	126	KRMHSLTMSFANSSI IKDHMLDIDRLVRFNLDSWS - - - KVLMEAKKITFELTVKQ	181
QY	193	LLGCEPÖLAGDSEÖQLVEAFEEMTRNULSPLIDVPFSGLYR - GMKARNLIHARIEÖNI	251
Db	182	LMSFDP - - - - GEWSESLRKEYLLVIEGFFSLPLPL - FSTYRKAIQARR - - - KVAEAL	231
QY	252	RAKICGLR - - ASEAGÖGCKDALÖLLI - - EHSWGERLDMÖALKÖSSTELLFGHETTAS	307
Db	232	TVVVMKRREEEGAEKKMDLALLAADDGFSDEIYDF - - - - LVALLVAGYEITST	285
QY	308	AATSLITYLGLVPHYLÖKVREELKSKGLCKSNÖDNKLDMEILEÖLK YIGCVIKETLRN	367
Db	286	IMTLAVKFLTETPLALAÖLKEE - - HEKIRAMKSDSYSLEWSDYKSMPTÖCVVNETLRVA	343
QY	368	PPVPGFRRVALKTFELNGYÖIPKGWNVIVSICDTHDVAEIFTNKEEFNPD RFSAPHPEDA	427
Db	344	NITGVFRRAMPTDVEIKGYKIPKGMKVVSSFRAVHLDPNHFKDARTFNPMRWÖSNSVTTG	403
QY	428	SRESFIPFGGLRSCVGEKEFAKILKIFVELARHCDWÖLLNGBPMTKTSPTVYPVDNLB	487
Db	404	PSNVFTPEGGPRLOPYELARVALSVFLHRLVTGFSW - - - - VPAEÖDKLVFFPTTRTQ	458
QY	488	ARFTHF 493	
Db	459	KRYPIF 464	

```

RESULT      8
US-08-991-677-4
; Sequence 4, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent L
; APPLICANT: Carraway, Daniel T
; APPLICANT: Smeltzer, Richard H
; TITLE OF INVENTION: Production of Syringyl lignin in Gymnosperms
; FILE REFERENCE: 50617
; CURRENT APPLICATION NUMBER: US/08/991,677A
; CURRENT FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: US 60/033,381
; EARLIER FILING DATE: 1996-12-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Liquidambar styraciflua
US-08-991-677-4

```

Query Match	13.3%;	Score 348.5;	DB 4;	Length 511;
Best Local Similarity	26.8%;	Pred. No. 1.4e-29;		
Matches 136; Conservative	80;	Mismatches 204;	Indels 87;	Gaps 22;

```
QY      14 FVLPLLLFLAAIKLMDLYCVSGRDRSCHLPRLPGTMGFFPFEGTLOAVLQRKFKLOMKRR   73
        |::||| | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      18 FIPLLLLLGLVSRL-----RQR--LPYPGPKGKLPVIGNMLMMDQLTHRGAKIAK   67

QY      74 KYGEIYKTHLFGRPTVRVMGADNVRRILIGDDRLYSVHWPASVFTILGGCLSNL-HDSS   132
        ::|:: : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      68 QYGGLFHLKMGFELHMVAVSTPDMARQVLOVDNIIFS-NRPATL-----AISLYTDRA   119

QY      133 -----HKORRKVIMRAFSSREALECY-----VPYTEVGSSLQEWMLSCG   171
        : || :|: ||| : : : | | | :|: :|: :|: :|: :|: :|: :|: :|: :|:
Db      120 DMAFAHYGPPWRQMRKLVCMKLFSSRKRAESMESVRDEVDASA VRVASNIGST---VNIG   175
```

QY	172	EKGLVYPVKRLMERIAMRILLGCEPOLAGDGDSEQOLVEAFEEMTR--NLFSLPIDVP	229
Db	176	E---LVEALTKNITYRAAFGTI-----SHEDQDEFVALIQEFSQLFGAFNIADFIIP	223
QY	230	F-----SGL-YRGMKARNLIHARIEONIRAKI--CGLRASEAGQGCKDALQLLIHSWER	281
Db	224	WLKWVPOGINVRLNKARGALDGFDIKIIDHIQKSGSKNSEVDTDMVDLLAFYGEAAKV	283
QY	282	GERLMO-ALK-----QSSTELFFGHETTASATSLITYLGlyPHVLQKYRELK-SK	333
Db	284	SESDDLQNSIKLTKNIKAIMDVMEFGETEVASAIEWAMTELMKSPEDLKRYOQELAVVY	343
QY	334	GLCKSNQDNKLDMLEILQKLYIGCVAIKETLRLLNPVPVPGGFVALKTFELNGYQIPIKGWN	393
Db	344	GL-----DRVEEKDFEKLTYLKCYLEKYLRLHPRIPLLHETAEDAEEVGYYIIPAKS R	397
QY	394	VLYSICDTHDVAEIFTNKEEFNPDRESAPHPED--ASRFSFIPEGGLRCSCYKEFAKIL	451
Db	398	VMINACAIGRDKNMSWADPDTFRPSRFPLKDGVDPDFKGNNEEFIIFGSGRRSCPQMÖGLYA	457
QY	452	LKIFTVELARHC-DWÖLLNGPPIMKTS	477
Db	458	LET-TVAHLHCFETWEL--PDGMKPS	480

```

RESULT          9
US-09-091-432-2
; Sequence 2, Application US/09091432
; Patent No. 5981837
;
; GENERAL INFORMATION:
;   APPLICANT: Chapple, Clint
;   TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition
;   FILE REFERENCE: 7024-325
;   CURRENT APPLICATION NUMBER: US/09/091,432
;   EARLIER FILING DATE: 1998-06-18
;   EARLIER APPLICATION NUMBER: PCT/US96/20094
;   EARLIER FILING DATE: 1996-12-19
;   EARLIER APPLICATION NUMBER: US 60/009,119
;   EARLIER FILING DATE: 1995-12-22
;   EARLIER APPLICATION NUMBER: US 60/013,388
;   EARLIER FILING DATE: 1996-03-14
;   NUMBER OF SEQ ID NOS: 3
;   SOFTWARE: Microsoft Word 2.0C
;   SEQ ID NO 2
;   LENGTH: 520
;   TYPE: PRT
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Sequence is deduced from DNA sequence of SEQ ID NO:1
;   Patent No. 5981837
;   US-09-091-432-2

```

Query Match	12.3%;	Score 322;	DB 2;	Length 520;
Best Local Similarity	25.0%;	Pred. No. 1.2e-26;		
Matches 129; Conservative	90;	Mismatches 198;	Indels 98;	Gaps 22;

```

QY 15 VLPLLLFLAIAIKLMDLYCVSGDRSCALPLRPGTMGFPEFGETLQWVLQRRKFLQMKRRK 74
   | : | : | : | | | | | | | | | | : |
Dd 22 VWSLFLFISFI-----TRRRRPYPYPPGPRGWPIGNNMLMDQLTHRGLANLARK 70
   | : | : | : | | | | | | | | | | : |
QY 75 YGFIYKTHLFGRPTVRVMGADNVRILLGDDRLVSVHWPASVRTLILSGCLSNL-HDSS- 132
   | | : : : | : | : | : | : | : | : | : | : |
Dd 71 YGICLHLMRGMFLHMYAVSSPEVARQVLQVDSVES-NRPATI-----AISLYTYDRAD 122
   | : | : | : | : | : | : | : | : | : |
QY 133 -----HKORKVIMRAFSEALLECYPV-----ITEEVGSSLEQWLSGGERGLL 176
   : | | : | : | : | : | : | : | : | : | : |
Dd 123 MAFAHYGPFWRQMRKKVCVMKVFSSRKRAESWASVRDEVDKMKVRSVSCNVGKPINVGEO--- 179
   | : | : | : | : | : | : | : | : | : |
QY 177 VYPEVKRLMFRIAMRILLGCPEQOLAGDGDSQQLVEAFEEEMTR--NLFSLPIDVPESGLY 234
   : : : : | | : | : | : | : | : | : | : | : |
Dd 180 IFALTRNITYRAFG--SACE-----KGODEFIRILQEFESKLFGAFNVADEPIPEGWI 230

```

```

QY 235 -----RGMKARN---LIHARIEQONIRAKICGLRASEAG-----QGCKDAQLLIENS 278
      | :||| | | : : : | | : | | :
Db 231 DPQGINRBLVKARNLDLDFIDDIIDEHMKKE-NQNAVDDGDVDTDMVDLLAFYSEEA 289
      | :||| | | : : : | | : | | :
QY 279 WERGERIDMQ-----ALKQSTELLFGSHETTASAATSLITYLGLYPHVLQVREEL 330
      | :|| | : | : :||| | | | | : | | : :|||
Db 290 KLVSEFADLQNSIKLTRONIKAIIMDVMEFGTETVASAIEWALTELLRSPEDLKRVQEL 349
      | :|| | : | : :||| | | | | : | | : :|||
QY 331 -KSKGLLCKSNQDNKLDMLEIQLKYIGCVIKETRLNRPVPVGGFRVALKTFELNGYQIP 389
      | :|| | : : : | :| | : :||| : :||| : :|||
Db 350 AEVVGTL-----DRVEESDIEKLTLYKCLIKETLRMHPRPIPLLHETAEDISIDGFFIP 403
      | :|| | : : : | :| | : :||| : :||| : :|||
QY 390 KGMNVIYSICDTHDVAEIFTNKEEFNPDFAHPED--ASRFSFIPEGGLRSVCVKEF 447
      | :||| | : : : | :| | | | | | | | :
Db 404 KKSRAVMINAFAIGRDPFTSWTDPDTRFPRSRFLPEGVPDFKGSNFEFIFPGSGRRSCPGMQ- 462
      | :||| | : : : | :| | | | | | | | :
QY 448 AKILKIFVELA----RHC-DWOLINGPPTMKT 477
      | :||| | | : :||| | | | |
Db 463 ----LGLVALDLAVAHILHCFWTKL---PDGMKPS 490
      | :||| | | : :||| | | | |

```

```

RESULT 10
US-08-948-564-16
; Sequence 16, Application US/08948564
; Patent No. 6121512
; GENERAL INFORMATION:
; APPLICANT: Siminszky, Balazs
; APPLICANT: Dewey, Ralph E.
; APPLICANT: Corbin, Frederick T.
; TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 6121512th Carolina
; COUNTRY: USA
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,564
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 5051-409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-948-564-16

```

Query Match	11.1%;	Score 289;	DB 3;	Length 576;
Best Local Similarity	23.7%;	Pred. No. 6.3e-23;		
Matches 128;	Conservative 86;	Mismatches 187;	Indels 138;	Gaps 22;

OY	41	ALPLPGT---	MGFPFEGETLQ	MLQRRKFLQ	MRKRRKYGF	LYKTHLFG	RPTVRVMG	ADN 96
Db	78	SMPiEAGVSD	LGRLPFLFSL	YDWFLF---	HGAVYKL-	AFGPKAFV	VSDDPI 125	

```

QY 97 VRRILLGDGR-----LVSVHWPASVRTILGSGCLSNLHDSHKKÖRKKVIMRAFSREAL 149
   | | : : : : : | : | : : : : |
Db 126 VARHILRENAFSYDKGLADILEP-----IMKGLIPADLD-TWKÖRRRVIAPAFHNSYL 179
   | | : : : : : : | | : : : : :
QY 150 ECYVPVITEEVGSLSLEQWLSGGERGLLVP-----EVRKLMFRIMRIL 193
   | | : : : : : : | | : : : : :
Db 180 EAMVKIFT-----TCSERTILKFKNLLEGEYDGDPSIELDLAEAFSSLAIDI 228
   | | : : : : : : | | : : : : :
QY 194 -LGCEPQLAGDGDSEQLVEA-----FEEMTRNLFSLP----- 225
   | | : : : : : | | : : : : :
Db 229 GLGVFNVDGFSVTKESPVIKAVYGTLEAEHRSTFYIPYWKIPLARWIVPRÖRKFFÖDDLK 288
   | | : : : : : : | | : : : : :
QY 226 -IDVPFSGLYRGMKARNLIHARIEONIRAKICGLRASEAGÖCKDA--LÖLLIEHSWEG 282
   | | : : : : : : | | : : : : :
Db 289 VINTCLDGLIRNAK-----ESRÖFEDVEK---LÖQRDY-LNLKASLLRFLVD--MRG 335
   | | : : : : : : | | : : : : :
QY 283 ERIDMÖALKÖSSTELLFGGHETTASATSLITYGLYPHYLÖKVREELKSKGLICKSNÖD 342
   | | : : : : : : | | : : : : :
Db 336 ADVDDRÖLRDDMLTMILAGHETTAVLTTWAVFLLAÖNPSKMKKAÖAEV--DLVLGTGR 392
   | | : : : : : : | | : : : : :
QY 343 NKLDMEILÖQLKYICGVIKETLRNDPVPGGFVALKT-----FELNGYÖIPKGMV 394
   | | : : : : : : | | : : : : :
Db 393 T---FESLKEIÖYIRLIVEALRLYPOPLLIRKSLKSDVLPFGHKGEKDGÖYAIÖAGTIV 449
   | | : : : : : : | | : : : : :
QY 395 IYSICDTHDVAEIFTNKEEFNPDF-----SAPHPEDA-SRFSF 432
   | | : : : : : : | | : : : : :
Db 450 FISVYNLHRSPYFWDRPDDEFERFLVÖNKNEIEGWAGLIDPSRGALYPNEYISDFAF 509
   | | : : : : : : | | : : : : :
QY 433 IPFGGGLRSCVKEFEAKILKIFTVELARHCDWÖLLNGPPTMK--TSPTVYPVDNLPAR 489
   | | : : : : : : | | : : : : :
Db 510 LPEGGGPRKCVGDÖFALMESTVALTMIONFDVELKGTPESEVELTÖGATIHITKNGMÖCR 568
   | | : : : : : : | | : : : : :

```

```

RESULT 11
US-08-948-564-4
; Sequence 4, Application US/08948564
; Patent No. 6121512
; GENERAL INFORMATION:
; APPLICANT: Siminszky, Balazs
; APPLICANT: Dewey, Ralph E.
; APPLICANT: Corbin, Frederick T.
; TITLE OF INVENTION: No. 6121512e1 Cytochrome P-450 Constructs and
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 6121512th Carolina
; COUNTRY: USA
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,564
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 5051-409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 510 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-08-948-564-4

Query Match	11.0%;	Score 288.5;	DB 3;	Length 510;
Best Local Similarity	24.5%;	Pred. No. 5.8e-23;		
Matches 119; Conservative	86;	Mismatches 220;	Indels 61;	Gaps 15;

[illegible]

RESULT 12
US-08-948-564-2
; Sequence 2, Application US/08948564
; Patent No. 6121512
; GENERAL INFORMATION:
; APPLICANT: Siminszky, Balazs
; APPLICANT: Dewey, Ralph E.
; APPLICANT: Corbin, Frederick T.
; TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 6121512th Carolina
; COUNTRY: USA
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,564
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.

REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 5051-409
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-854-1400
TELEFAX: 919-854-1401
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-564-2

Query Match 10.4%; Score 271.5; DB 3; Length 513;
Best Local Similarity 25.3%; Pred. No. 4.4e-21;
Matches 132; Conservative 71; Mismatches 214; Indels 105; Gaps 23;

QY 5 ALLASAL-----CTFVLPPLLFLAIIKIMDLVCSGRDRSCALPLPPG 47
Db 2 ALLSSVLKQLPHELSTHYLTVFCIFLLQLI-----RRNKYNLPPS 45
QY 48 TMGPFPEGETLQMVQRRKFLQMKRRKYGFIYKTHLEGRPTVRVMGADNVRRIILGDDR 106
Db 46 PPKPIIIGNLHQLGTLPHRSFHALS-HKYGPLMLQLQIPTLVSSADVAIREIKTHDV 104
QY 107 LVSVHWPASVRTILGSGCLSNLHDSHKQKKVIMRAFSREALE-CYVPV----- 155
Db 105 VFSNRROPTAAKIFGYGC-----KDVAFVYRREWRQKIKTCKVEIMSLKKVRLFH 155
QY 156 -ITEEVGSSLEQWL--SCG-ERGLLVPEVKRLMR---IAMRILLG--CEPOLAGDGS 206
Db 156 SIRQEVNTELVEAIGACGSEKPCVNLTQM--LMAASNDIVSRVGLGRKCDACGSGSS 213
QY 207 EQQLVEAFEEEMTRNL-----FSLPIDVP-----FSGLYRGMKARNL-IHARTEQNIR 252
Db 214 -----SFAALGRKIMRLLSAFSVGDFPPLSGWVDYLTGLIPEMKTFFLAVADLDEVIA 267
QY 253 AKICGLRASEAGQCKDALQLLIEHSWERGE---RLDMQALKOSSTELLFGGHETTASAA 309
Db 268 EHSSNKKND-----DFLGIQLQLQ-ECGRLDFOQLDRDLKAILVDMITIGSDPTSTTL 320
QY 310 TSLITYLGLPHYLQKVRLELKSGLLCKSNQDNKIDMEILEQLKYGIVIKETRLRNP 369
Db 321 EMTFAEFLRNPNTMKKAQEEVRR--VVGINSKAVLDENCVMQNMNYLKCVAKETRLRHP 377
QY 370 VPGGFRVALKT---FELNGYQIPKGMNVISICDTHDVAEIFTNKEEFNPDRSAPHPE- 425
Db 378 LP--LLIARETSSSVKLRGYDIPAKTMVFINAMAIQRPDLMDPDEEFIPEREFTSQVDL 435
QY 426 DASRFSFIPEGGGRSCVSGKEFAKILKIFVELARHCDWQL 467
Db 436 NGQDFOLIPIFGIGRRGCPAMSFGLASTEYVLANLLYWNWNN 477

RESULT 13
US-08-313-075A-50
Sequence 50, Application US/08313075A
Patent No. 5639870
GENERAL INFORMATION:
APPLICANT: Holton, Timothy A.
APPLICANT: Cornish, Edwina C.
APPLICANT: Tanaka, Yoshikazu
TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID
TITLE OF INVENTION: PATHWAY ENZYMES AND USES THEREFOR
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: U.S.A.
ZIP: 11530

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313, 075A
FILING DATE: 30-NOV-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL 1538/92
FILING DATE: 27-MAR-1992
APPLICATION DATA:
APPLICATION NUMBER: AU PL 6698/93
FILING DATE: 07-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU93/00127
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Digilio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 9433
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-313-075A-50

Query Match 10.1%; Score 264.5; DB 1; Length 496;
Best Local Similarity 24.4%; Pred. No. 2.5e-20;
Matches 126; Conservative 66; Mismatches 205; Indels 119; Gaps 22;

QY 20 LFLAIIKLMDLYCVSG---RDRSCALPLPPGTMGPFPEGETLQMVQRRKFLQMKRRKY 75
Db 6 ILLGLFTW--FLVNGLSLRRRKISKKLPPGPFPLIIGNLHLGNHPHKSIAQLAKIH 63
QY 76 GFYKTHLEGRPTVRVMGADNVRRIILGDD-----RLV-----SVHW-PA--- 114
Db 64 GPIMNLKLGQNTVYISSSVAREVLQKODLTFSNRFVPDVVHVRNHSDFSVWLLPVNSR 123
QY 115 --SVRTILSGGCLS--NLHDSHKQKKVIMRAFSREALECYVPVITTEVGSSLEQWLSC 170
Db 124 WTLRKIMNSSIFSCKNIDGNQHLRSKV-----QEL---IDYCQKC 162
QY 171 GERGLLVPEVKRLMRIAMRILIGC--EPQLAGD-GDSEQQLVEAFEEEMTRNL----- 221
Db 163 AKNGEAV--DIGRAFTGTLNLSNTIFSKDLTFNPFSDSAKE---FKELVWNIMVEAGK 216
QY 222 -----FSLPID-----VPFSGLYRG-----MKARNLIHARIEQNIRAKIC 256
Db 217 PNLVDYFPFLKIDPOGIRKRTNNTFTKFLGLISGLIDRLKERNL---ROMANI----- 268
QY 257 GLRASEAGQCKDALQLLIEHSWERGERLDMQALKOSSTELLFGGHETTASATSLITYL 316
Db 269 -----DYDALNLNISQENPEEIDRNOIEQLCLDLFAAGTDTTSTNLEWAMAEI 316
QY 317 GLPHYLQKVRLELKSGLLCKSNQDNKIDMEILEQLKYGIVIKETRLRNPVPGGF-R 375
Db 317 LQNPHTLOKAQELAAQ--VIGKGQVEEAD---VGRLLPYLRCAIVKETLRIHPAAPLLIPR 371
QY 376 VALKTEFLNGYQIPKGMNVISICDTHDVAEIFTNKEEFNPDRFSAPHEDASR-FSFI 434
Db 372 KVEDVELSTYIIPKDSQVLNVWMAIGRNSDLMENPLVEKPERFWESEIDIRGRDFELLIP 431
QY 435 FGGGLRSCVSGKEFAKILKIFVELARHCDWQLNG 470

Db 432 FGAGRRICPGLPLAMRMIPVALGSLNSFNWKLYGG 467

RESULT 14

US-09-292-768-68
; Sequence 68, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; FILE REFERENCE: wsur13463
; CURRENT APPLICATION NUMBER: US/09/292,768
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 68
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-292-768-68

Query Match 9.7%; Score 252.5; DB 4; Length 500;
Best Local Similarity 21.7%; Pred. No. 5.3e-19;
Matches 114; Conservative 86; Mismatches 217; Indels 109; Gaps 18;

QY 7 LASALCTFVLPPLLFLAIAIKLMDLYCVSGRDRSCALPLPGTMGFPFEGETLQMYLQR-- 64
::||: | : | || | |||| | | : :
Db 5 ISSAIIILVVTYTIISLLIKQW-----RKPQENLPPGPKPLIGH-LHLLMGKLP 56
:
QY 65 RKFLQMKRRKYGFYKTHL-----FGRPTVRVMGADN 96
: | : || : | : : : : : : : : :
Db 57 QHALASVAKQYGPVAHVQLGEVFSVLSREATKEAMKLVDPACADRFESIGTKIMWYDN 116
:
QY 97 VRRILLGDDRL--VSVHWPASVRTILGSGCLSNLHDSHKQKKVIMRAFSREALCYV 153
||: | : |||| : || : | : :
Db 117 -----DIIIFSPYSVHW-----RQMRKICVSELLSARNVRSFG 149
:
QY 154 PVITEEVGSSLEQWLSGGERGLLVPEVKRMFRIA-MRILLGCEPOLAGDGDSEQQLVE 212
: ||: | : | : | ||| : | : : ||
Db 150 FIRQDEVSRLLGHLRSSAAG----EAVDLTERIATLTCSIICRAAFGSVIRDHIELV- 203
:
QY 213 AFEEMTRNLFSLPIDVPFSGLYRGMKARNLI-----HARIEQNIRAKICGLRASE 262
| : : | : : : | : | : : : : : : :
Db 204 ---ELVKDALSMASGFELADMFPSSKLNLLCWNKSKLMRMRRRVDAILLEAIVEEHKLLK 260
:
QY 263 AGQ-GCKDALQLLIEHSWGERLDM-QALKQSTELLFGGHETTASATSLITYGLY 319
: | : | : | : : : : : : : : : : : :
Db 261 SGEEGGEDIIDVLFMRQKDSQIKVPITTKAIFIDTFSAGTETTSITTLWMAELMRN 320
:
QY 320 PHVLQKVRRELKSKGLLCKSNQDNKLDMEILEQLKYIGCVIKETLRLNPPVGGFRVALK 379
| : | : | : : : | : : : ||| : : : : : :
Db 321 PEVMAKAQAEYRA--ALKGKTD-WDVDDVQELKYMKSVMKETMRMHPPIPLIPRSORE 375
:
QY 380 TFEELNGYQIPKGMNVISICDTHDVAEIFTNKEEFNDRFSAPHPEDASR-----FSFI 433
| : ||| | : : : : | | | : : : : :
Db 376 ECEVNGYTIIPNKARIMINWMSGRNPPLYWEKPETFWPERF-----DQVSRDFMGNDFEFI 430
:
QY 434 PFGGGLRSCVGEKEFAKILKIETVELARHCDWQLNG--PPTMKT 477
||| | | | : : : : | | | | | :
Db 431 PFGAGRRICPGLNFGLANVEVPLAQLLYHFDWKLAEGMNPSDMDMS 476

RESULT 15
US-09-292-768-70
; Sequence 70, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:

; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; FILE REFERENCE: wsur13463
; CURRENT APPLICATION NUMBER: US/09/292,768
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-292-768-70

Query Match 9.7%; Score 252.5; DB 4; Length 500;
Best Local Similarity 21.7%; Pred. No. 5.3e-19;
Matches 114; Conservative 85; Mismatches 218; Indels 109; Gaps 18;

QY 7 LASALCTFVLPPLLFLAIAIKLMDLYCVSGRDRSCALPLPGTMGFPFEGETLQMYLQR-- 64
::||: | : | || | |||| | | : :
Db 5 ISSAIIILVVTYTIISLLIKQW-----RKPQENLPPGPKPLIGH-LHLLMGKLP 56
:
QY 65 RKFLQMKRRKYGFYKTHL-----FGRPTVRVMGADN 96
: | : || : | : : : : : : : : :
Db 57 QHALASVAKQYGPVAHVQLGEVFSVLSREATKEAMKLVDPACADRFESIGTKIMWYDN 116
:
QY 97 VRRILLGDDRL--VSVHWPASVRTILGSGCLSNLHDSHKQKKVIMRAFSREALCYV 153
||: | : |||| : || : | : :
Db 117 -----DIIIFSPYSVHW-----RQMRKICVSELLSARNVRSFG 149
:
QY 154 PVITEEVGSSLEQWLSGGERGLLVPEVKRMFRIA-MRILLGCEPOLAGDGDSEQQLVE 212
: ||: | : | : | ||| : | : : ||
Db 150 FIRQDEVSRLLGHLRSSAAG----EAVDLTERIATLTCSIICRAAFGSVIRDHIELV- 203
:
QY 213 AFEEMTRNLFSLPIDVPFSGLYRGMKARNLI-----HARIEQNIRAKICGLRASE 262
| : : | : : : | : | : : : : : : :
Db 204 ---ELVKDALSMASGFELADMFPSSKLNLLCWNKSKLMRMRRRVDAILLEAIVEEHKLLK 260
:
QY 263 AGQ-GCKDALQLLIEHSWGERLDM-QALKQSTELLFGGHETTASATSLITYGLY 319
: | : | : | : : : | : | : : ||| : : : : : :
Db 261 SGEEGGEDIIDVLFMRQKDSQIKVPITTKAIFIDTFSAGTETTSITTLWMAELMRN 320
:
QY 320 PHVLQKVRRELKSKGLLCKSNQDNKLDMEILEQLKYIGCVIKETLRLNPPVGGFRVALK 379
| : | : | : : : | : : : ||| : : : : : :
Db 321 PEVMAKAQAEYRA--ALKGKTD-WDVDDVQELKYMKSVMKETMRMHPPIPLIPRSORE 375
:
QY 380 TFEELNGYQIPKGMNVISICDTHDVAEIFTNKEEFNDRFSAPHPEDASR-----FSFI 433
| : ||| | : : : : | | | : : : : :
Db 376 ECEVNGYTIIPNKARIMINWMSGRNPPLYWEKPETFWPERF-----DQVSRDFMGNDFEFI 430
:
QY 434 PFGGGLRSCVGEKEFAKILKIETVELARHCDWQLNG--PPTMKT 477
||| | | | : : : : | | | | | :
Db 431 PFGAGRRICPGLNFGLANVEVPLAQLLYHFDWKLAEGMNPSDMDMS 476

Search completed: November 6, 2001, 13:27:05
Job time: 9918 sec

```
QY      341 QDNKLDMEILEQLKYIGCVIKETRLRNPVPVGGFRVALKTFFELNGYOIPKGWNVIYSICD 400
        :| |:::|: |:| ||||| :::|: |||| | | |
Db      289 LSOELTAETIKKMPYLQVLRLRPVGGGFRLEIQDCQFQGFIHPKGMWLSYSIQ 348
QY      401 THDVAEIFTNKEEFNDRF---SAPHPEDASRSFIFGGGLRSCVCGKEFAKILKITFT 456
        || :::|:|:|:| || | | :||||||| |:|||:::|:|
Db      349 THADPDLYPDPPEKFDERFTPDGSATHNP--FAHVPFGGGLRECLGKEFAELEMKLFA 405
QY      457 VELARHCDWLNGPP-TMKTSPTVYPVDNL 486
        | : ||||| : :|: | |||
Db      406 TRLIQGFWDWTLLPGQNLELVTPSPRPKDNL 436
```

RESULT 2
A84859
probable cytochrome P450 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 23-Mar-2001
C:Accession: A84859
R:Llin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nlerman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: A84859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-485 <STO>
A:Cross-references: GB:AE002093; NID:g4512670; PIDN:AAD21724.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g42850
A:Map position: 2
C:Superfamily: Synecocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F;432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	19.98;	Score 519.5;	DB 2;	Length 485;
Best Local Similarity	28.38;	Pred. No. 2.9e-32;		
Matches 132;	Conservative 86;	Mismatches 201;	Indels 47;	Gaps 10;
QY 44	LPEGTMGPEPFEGETLQMLQRRK-----FLQMKRRKYGYFKTTLFGRPTRVRMGADNV	97		
Db 45	LLPGEGLPWIGETMDFYKQAKSNRVFEDEVNPRIIKHGNIKFTRIMSGPTIVNGAEAN	104		
QY 98	RRILGDDRLVSVHWPASVRTILGSGCLSNLHDSHKQKKVIMRARSREALCYVPVIT	157		
Db 105	RLIISNEFSLVSSWSPSSSVQLMGNMCIMAKQGEKHRVLRGIANSISYIGLESILPKLC	164		
QY 158	EYVGSSLE-QWLSGGERGLVYPEVKRIMFRIAMRILLGCEPQLAGDGDSEQLVEAFEE	216		
Db 165	DTVKFHHETWR--GKEIISLYRSARVLTFTVVECLYGIKVEIG-----MLEVFER	214		
QY 217	MTRNLFSLPIDVPFSGLYRGMKARMLIHARIEQNIIRAKICGLRASEAGQCKDALQLLIE	276		
Db 215	VLEGVFALPVEFPCKSFARAKRADEIETFLVGKVRK---RREMEKEGAEPNTTLF-	269		
QY 277	HSWGERGLDMQALKQSSSTE-----LIFGHETTASAATSLITYLGLYPHVLQKVR	327		
Db 270	-----SRLVEELIKGVITFEDEVVDMVLLVFAAHDTSYAMSMTEFKMLAQHP----TCR	319		
QY 328	EEKSKGLLCKSN--QDNKLDMEILIEQLKYIGCVIKETTLRLNDPVGGERVALKTFELNG	385		
Db 320	DTLLQEHQAQKANKGEGEYLLVEDVKKMKYSWQVVRETMRLSPPIFGSFRKAVADIDYGG	379		
QY 386	YQIPKGWNVIYSICDTHVAETFTNKEEFNPDRFSAPHPEDASRFSFIPIFGGLRSCVVK	445		
Db 380	YTIPIKGMKILMTTGYGTHYNPETIQDPMSEFDRFKP---IQAYTYLLPFGGGPRLCAGH	435		
QY 446	EFAKILLKFTFVELARHCDWQLNGPPTMKTSPTVVYPVDNLPARFT	491		

Db 436 QLAKISILVEMHVVVTGFDWSLVYPDETISMDDLPPSLGMPKIS 481

RESULT 3

T04444

cytochrome P450 - Arabidopsis thaliana

N;Alternate names: protein T18B16.200; protein T5K18.10

N;Contains: oxidoreductase (EC 1.-.-.-)

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2000

C;Accession: T04444; T05806

R;Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Anson, W.; Bancroft, I.; Mewes, submitted to the Protein Sequence Database, April 1998

A;Reference number: Z15359

A;Accession: T04444

A;Molecule type: DNA

A;Residues: 1-457 <BEV>

A;Cross-references: EMBL:AL021687

A;Experimental source: cultivar Columbia; BAC clone T18B16

R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; submitted to the Protein Sequence Database, April 1998

A;Reference number: Z15453

A;Accession: T05806

A;Molecule type: DNA

A;Residues: 131-457 <BEW>

A;Cross-references: EMBL:AL022580

A;Experimental source: cultivar Columbia; BAC clone T5K18

C;Genetics:

A;Map position: 4

A;Introns: 67/2; 173/3; 302/3; 358/1; 393/3

A;Note: T18B16.200; T5K18.10

C;Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

C;Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreducta

F;272-433/Domain: cytochrome P450 homology <P45>

[illegible]

T07859
cytochrome P450 homology - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 16-Feb-2001
C:Accession: T07859
R:Bishop, G.J.; Harrison, K.; Jones, J.D.
Plant Cell 8, 959-969, 1996
A:Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes the
A:Reference number: Z16181; MUID:96266705
A:Accession: T07859
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-464 <BIS>
A:Cross-references: EMBL:U54770; NID:g1421740; PIDN:AAB17070.1; PID:g1421741
A:Experimental source: strain GCR758
C:Genetics:
A:Gene: dwarf
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
F:273-436/Domain: cytochrome P450 homology <P45>

Query Match	17.0%;	Score 444;	DB 2;	Length 464;
Best Local Similarity	27.48;	Pred. No. 1.8e-26;		
Matches 128; Conservative	94;	Mismatches 203;	Indels 42;	Gaps 12;

```

QY 18 LLEFLAIAKLMWLDLYCVSGRDRSCALDLPBPTMGEPPEFGETLQMWLQRRKEFLQMKRRKYGE 77
   | : | : | : : : | | | | | : : : | : | : |
Db 14 LCIFCTALLRWNOVKYNOKN-----LPPGTMGWPLFGETTEFLKLGSPFMKNQRRARYGS 67

QY 78 IYKTHLEGRPTVRVMGADNVRRIILGDDRLVSVHWPASVRTILGSGCLSNLHDSHKQOK 137
   : | : | : | : : | : : : : | : | : : | : | :
Db 68 FFKSHILGCPITVSMDELNRYILVNEAKGLVPGYPQSMIDILGKCINIAVNGSAHKMYR 127

QY 138 KYIMRAFS---REALCYVPVITEEVGSSLEQWLSCGERGLVYPEYKRIMERIAMRIL 193
   : : | : | : | : | : | : : : : : : : : : :
Db 128 GALLSLISPTMIRDQL--LPKIDEMRSHLTNW--DNKVIDIQEKTNKAFLSSLKQI 181

QY 194 LGCE-PQLAGDGDSEQOLVEAFEEMTRNLFSLPIDVPFSGLYRGMKAR---NLIHARIE 248
   | | | | : : : : | : | : | : : : | : | : |
Db 182 AGIESTSLA-----QEFMSEFFNLVGLTSLPINLPNTNYHRGFQARKIIVNLLRTLIE 235

QY 249 QNIRAKICGLRASEAGQGCKDALQLLIHSWGERLDMQALKQOSTELLEFGHETTASA 308
   : : | : | : | : | : | : : | : | : | : : :
Db 236 ER-----RASKETQ--HDMGLYLMNEBATTRFKLTDDDEMIDLIT-IYSGETVSTT 284

QY 309 ATSLITYLGLYPHYLQKVREELKSKGLCKSNQDNKLDMEILLEQLKYIGCVAIKETRLNP 368
   : : | : | : | : | : | : : : : : : : : :
Db 285 SMAAVKYLHDHPKVLBELRKE--HMAIREKKKPEDPIDYNDYRSMRFTRAVILETSRLAT 342

QY 369 PVEGCFRVALKTFELNGYQIPKGNVYISICDTHDVAEIFTNKEEFNPDRESAPHEDAS 428
   | | | : : | : | | | : : : : : : : : : :
Db 343 IVNGVLRKTTQDMEINGYIIPKGWRIYVYTRELNYDPRLYPDPYSFNPWRMMDKSLHQN 402

QY 429 RFSFLPFGGGLRSCVKGKEFAKILKIFVELARHCDWQLNGPPTMK 475
   | : | | | | | | : | : | : : | : : | : |
Db 403 --SFLVEGGGTROCPGKELGVAEISTFLHYFTKYRWEIEGGDKLMK 447

```

RESULT 7
T46143
steroid 22-alpha-hydroxylase (DWF4) - Arabidopsis thaliana
N;Alternate names: protein T3A5.40
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 16-Feb-2001
C;Accession: T46143
R;Bloembergen, H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueler, C.; Quetier, F.; Sala
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23024
A;Accession: T46143
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-513 <BLO>
A;Cross-references: EMBL:AL132979

A; Experimental source: cultivar Columbia; BAC clone T3A5
C; Genetics:
A; Map position: 3
A; Introns: 74/2; 182/3; 233/3; 338/3; 369/3; 396/1; 432/3
A; Note: T3A5.40
C; Superfamily: *Synechocystis* cytochrome P450 slr0574; cytochrome P450 homology
C; Keywords: heme; iron; metalloprotein
F; 308-484/Domain: cytochrome P450 homology <P45>
F; 462/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	16.8%;	Score 439;	DB 2;	Length 513;
Best Local Similarity	27.0%;	Pred. No. 5e-26;		
Matches 142;	Conservative	85;	Mismatches 226;	Indels 72;
				Gaps 15;

```

QY 3 LPALLASALCTFVLPLLLFLAIAIKLMDLYCVSGRBSICALPLPBTGMGFPEFGELL-----58
   || || : | |||| : | || | |||| : ||| : || ||| :
Db 10 LPLLLPSL---LSLLFLILLK-----RRNRKTRFNLPBGKSGWPFLEGTIGYLK 57
QY 59 -QWVLQRRKFLQMKRRKYGFYKTHLEGRPTVRVMGADNVRILLGDRLVSYHMPASVR 117
   | : | ||| : ||| : | || : || : || : || :
Db 58 PYTATTLGDFMQOHVSKYKTYRSNLFGEPTIVSADAGLNRFILQNEGRLLFECSYPRSIG 117
QY 118 TLISGCLSNLHDSHKQRRKVMIRAFSREALCYVPVITEEVGSS---LEQWISGGER 173
   ||| : | : : : : | | : : : || : |
Db 118 GILGKWSMLVLVGDMHRDMRSISLNLSHARLR--TLLKDVHRHTLFVLDSWQO--NS 172
QY 174 GLUYPEVKRIMFRIAMRILLGCEPQLAGDSE-QOLVEAFEEMTNRLFSLPIDVPFSG 232
   | : | : : : : | : | : : : : : | : || : |
Db 173 IFSAQDEAKKFTFNIMAKHIMSMDP---GEEETEQLKRYVTFMKGVSAPLNPGTA 227
QY 233 IYRGMKARNLIHARIEONIRAKICGLRASEAGQCKDALOLLIHSWE-----RGERL 285
   :: : : | | : : : : : : : : | | : | : |
Db 228 YHKAQSRATILKFIERKMEERKLIDIKEED---QEEEVKTEDEAFEMSKSDHVRKORT 282
QY 286 DMQ---ALKQS--STE-----LLEFGHETTASATSLITYLGLYPHVLQKVBELK 331
   | || | || | || || || || : | : | : : |||
Db 283 DDDLGLWVLKHSNLSREQILDLILSLLFAGHETSSVAIALAIFLQACPKAVEELREHLL 342
QY 332 SKGLLCKSNODNKLDMELLQQLKYIGCVIKETLRLNRPVGGFRVALKTFELNGYQIPKG 391
   | : : : : : : : || || || | || | || |
Db 343 EIARAKKELGESELNWDYKKNDFYQCVINETLRLGNVVRFLHRKALKDVRYKGYDIPSG 402
QY 392 WNVYISICDTHDVAEIFTNKEEFNPDREFASPHPEDASRFS-----FIFPGGGLRSC 442
   | : | : | : : ||| : : : | : : |||| | |
Db 403 WKVLPVISAVHLDNSRYDPENLEFPWRWQOQNNGASSSGSGSFSTWGNNYMPFGGGRLLC 462
QY 443 VGKEFAKILLKIFTVELARHCDMOLLNGPPTMKTSPVTVPVNDLP 487
   | | || : : | : : || : | : : | : |
Db 463 AGSEIAKLEMAVFIHHLVLKFNWELAE-----DDKPFAPFVDFP 502

```

RESULT 8
B84733
probable cytochrome P450 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: B84733
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: B84733
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-489 <STO>
A;Cross-references: GB:AE002093; NID:g3831452; PIDN:AAC69934.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g32440
A;Map position: 2
C;Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Db 170 EFEETIKGLICIPKPEGTRLYKSLKAKERLIKMKVKEEROVAMTTTSPANDVYDVL 229
QY 273 LLIHESWGERLDMQALKOSSTELLFGHETTASATSLITYLGLYPHYLOKVEELKS 332
Db 230 RDGDSEKQSPDSFVSGK--IVEMMIPGEETMPTAMTLAVKFLSDNPVALAKLVEE--- 284
QY 333 KGLLCKSNQDNKLDMEILEQKLYIGC-----VIKETLRLNPPVPGGFVALKTFEELN 384
Db 285 -----NMEMKRRLKEIGEYKWTDMYMSLSTQYNINETLRMANIINGVWRKALKDVEIK 338
QY 385 GYQIPKGMNVIYSICDTHDVAEIFTNKEEFNPDRESAPHEDASRFSFIPEGGLRSCVG 444
Db 339 GYLIPKGCVLASFISYHDEIDYDNPYQFDPWRMDRINGSANSSICFTPFGGGQLCPG 398
QY 445 KEFAKILLKIFTVELARHCDWQLLNGPPTMTKSPVYEPVDNLPAR 489
Db 399 LELSKLEISIFLHHLVTRYSWTA--EEDIVSFPTVKMKRRLPIR 441

RESULT 11
D70649
cytochrome P450 Rv3059 - Mycobacterium tuberculosis (strain H37RV)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: D70649
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987
A: Accession: D70649
A: Status: nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-492 <COL>
A: Cross-references: GB:Z83866; GB:AL123456; NID:g3261691; PIDN:CAB06263.1; PID:g1781154
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: RV3059
C: Superfamily: Mycobacterium cytochrome P450 Rv3059; cytochrome P450 homology
C: Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F: 300-461/Domain: cytochrome P450 homology <P45>
F: 439/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 15.3%; Score 400; DB 1; Length 492;
Best Local Similarity 28.2%; Pred. No. 4.7e-23;
Matches 137; Conservative 81; Mismatches 192; Indels 76; Gaps 21;
QY 37 DRSCALPLPPT-----MGPFEGETLQWVLORRKFLQMKRRKYGYIKYTH---LFG 85
Db 42 EKCLAEP-PPSGLRPVVGDAGLPILGHMIEMLRGPDYIM-----FLYKTKGPVVEG 93
QY 86 ----RPTVRVMGADNVRRIILLGDDRLVSVH-WPASVRTILGSGCLSNLHDSHQRKKVI 140
Db 94 DSAVLPGVAALGPDAQVIYSNRNKDYSQGWVPYIGPFFHRG-LMLLDFEEHMFHRRIM 152
QY 141 MRAFSREALCEYVPVITEEVGSSL-EQWLSGGERGLLVPEVKRLMFRIAMRILGCEPQ 199
Db 153 QEAFVRSRLAGYLEQMDRVVSRVADDWV-VNDARFLVYPAMKALTLDIASMVFMCHEP- 210
QY 200 LAGDGDEQQLVEAFEEEMTRN-----LPSLPIDVPFSGLYRGMKARNLIH---ARIEQN 250
Db 211 -GTDHELVTYVKNKAFTITTRAGNAVIRTSVP--PFT-WWRGLRARELLENYFTARVKE- 264
QY 251 IRAKICGLRASEAGCGKDALQLLIHESWGERLDMQALKOSSTELLFGGHETTASAT 310
Db 265 -----RREASG---NDLLTVLCQTEDDDGNRPSDADIVNHMIFLMAAHDTSTSTAT 313
QY 311 SLITYLGLYPHYLOKVEELKSKGLLCKSNQDNKLDMEILEQKLYIGCVIKETLRLNPPV 370

Db 314 TMAVQLAAHPPEWQORCRDESDRHG-----DGPLDIESLEQLESLDLVMNESIRLVTPV 366
QY 371 PGGFVALKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNPDRESAPHPE-DASR 429
Db 367 QWAMRQTVRDELLGYLLPKGTNVIAYPGMNHRLLPELWTDPLTFDPERFTEPRNEHKRHR 426
QY 430 FSPFPEGGLRSCVGKKEFAKT-----LKIFTVELARHCDWQLLNGPPTMTKSPVY 481
Db 427 YAFTFEGGVHCKIGWVFQLEIKTILHRLRLRYRLELSRP-DYQ----PRWDYSAMPI 480
QY 482 PYDNLP 487
Db 481 PMDGM 486

RESULT 12
T02263
cytochrome P450 DWARF3 - maize
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Zea mays (maize)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: T02263
R: Winkler, R.G.; Helentjaris, T. Plant Cell 7, 1307-1317, 1995
A: Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibbe
A: Reference number: Z14648; MUID: 96004534
A: Accession: T02263
A: Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: mRNA
A: Residues: 1-519 <WIN>
A: Cross-references: EMBL:U32579; NID:g987266; PIDN:AAC49067.1; PID:g987267
A: Experimental source: strain B73
C: Genetics:
A: Gene: dwarf3
C: Function:
A: Description: involved in an early step in gibberellin biosynthesis
A: Pathway: gibberellin biosynthesis
C: Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
C: Keywords: oxidoreductase
F: 325-488/Domain: cytochrome P450 homology <P45>

Query Match 14.9%; Score 388; DB 1; Length 519;
Best Local Similarity 27.2%; Pred. No. 4.3e-22;
Matches 126; Conservative 86; Mismatches 216; Indels 36; Gaps 14;
QY 44 LPBGTMGPFEGETLQWVLORRK----FLQMKRRKYG--FIYKTHLFGRPVRYMGADN 96
Db 72 LPPGEMGWPLVGGWMAFLRAFKSGKPDAFIASFVRRTGTGYRSFMFSSPTVLYTTAEG 131
QY 97 VRRILLGDDRLVSVHWPASVRTILGSGCLSNLHDSHQRKKVIMRAFSR-EALECYVPV 155
Db 132 CKOYLMDDDAFVT-GWPRATVALVGPFRSFVAMPYDEHRRIRKLTAPINGFDALTGYLPF 190
QY 156 ITEEVGSSLQWLSGGERGLLV-PEVKRLMFRIAMRILGCEPQLAGDGDEQQLVEAF 214
Db 191 IDRTVYSSLRAWADHG--GSVEFTELRRMTFKIIVQIFLG-----GADQATYRALERSY 243
QY 215 EEMTRNLPSLPIDVPFSGIYRG-MKARNLIHARIEQNIRAKICGLRASEAGCGKDALQL 273
Db 244 TELNYGRAMAINLP-GPAYRGALRARRLVA-VLQGVLDERRARAKVSGGVDMMDR 301
QY 274 LIEHSWGERLDMQALKOSSTELLFGGHETTASATSLITYLGLYPHYLOKVEELKSK 333
Db 302 LIEAQDERGRLHDDDELIDVLYMLNAGHESSGHTMWTATVFLQENPDMFARAKAEQEI 361
QY 334 GLLCKSNQDNKLDMEILEQKLYIGCVIKETLRLNPPVPGGFVALKTFELNGYQIPKGMN 393
Db 362 MRSIPSSQ-RGLTLRDRFRMEYLSQVIDETLRLVNTSVFSFQATRDVFNGLYIPKWK 420
QY 394 VIYSICDTHDVAEIFTNKEEFNPDRESAPHPEASRFSFIPEGGLRSCVGKKEFAKILLK 453

Db 421 VOLWYRSVHMDPOVYPDPTEKFDPSRWEHSPRAG---TFLAFGLGARLCPGNDLAKLEIS 477

QY 454 IFT-----VELARHCDWQLNGPPTMKTSPTIYPVDNLPARFT 491

Db 478 VFLHHEFLGLYKTLAR-----TNPRCRVRYLPHRPYDNCCLKIT 515

RESULT 13

F70791

Cytochrome P450 RV3685C - Mycobacterium tuberculosis (strain H37RV)

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Mycobacterium tuberculosis

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C/Accession: F70791

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A/Accession: F70791

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-476 <COL>

A/Cross-references: GB:AL022121; GB:AL123456; NID:g3261559; PIDN:CA18007.1; PID:g296010

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: RV3685C

C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase

F;298-444/Domain: cytochrome P450 homology <P45>

F;422/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.0%; Score 338.5; DB 1; Length 476;

Best Local Similarity 25.2%; Pred. No. 2.5e-18;

Matches 102; Conservative 76; Mismatches 178; Indels 49; Gaps 10;

QY 65 RKLQMKRRKYGFIYKTHLFGPRTVRVMGADNVR-----ILGDDRLVSVHWPASV 116

Db 68 RRLRLIRR-YGPIMTPIILSLGDVAIVSDALAKEVFTAPTDVLLGGEV---GPAA- 121

QY 117 RTILSGCLSNLHDSHKOKKVIIMRAFSEALECYVPVITEEVGSSLEQWLSGGERGL 176

Db 122 -AIYSGSGMFVQEEPEHLRRKLLTPPLHGAALDRYVPIENSTRAAMHTWPV--DRPFA 178

QY 177 VYPEVKRLMFIAMRILLCCEPQLAGDSEQLVEAFEEEMTRNLFSLPIDVPFS-----G 232

Db 179 MLTVARSIMLDVIKVFQD-----DPEVRRILGRPFERLNLGVSEQLTVRYALRRLG 233

QY 233 LYRGMKARNLIHARIEQNIIRAKICGLRASEAGGCKDALQOLLIEHSWGERGLDMQALKQ 292

Db 234 ALRWPARARANTEIDDVWALIAQRADPRLGERHDVLSLVSARGESGEQLSDSEIRD 293

QY 293 STELLFGHETTASATSLITYLGLYPHYLVOKVREELKSKGLCKSNQDNKLDMEILEQ 352

Db 294 DLITLVLAGHETTATTLAFAFDLLIHPDALRRVRAEAVGG----- 335

QY 353 LKYGCVIKETLRLNPPVPGFRVALKTFELNGYQIPKGMVYISICDTHDVAEIFTNKE 412

Db 336 EAFITAVINETLRVPPAPLTAARVAAQPLTIGGYRVEAGTRIIVVHIIAIRSAEVEYHPH 395

QY 413 EFNDPDSAPHPEDASRFSFTPEGGLRSCVKGKEFAKILLKIFTV 457

Db 396 EFRPERFLGTRPQ---TYAWVDFGGGVKRCIGANFS--MRELITV 435

RESULT 14

H96759

probable steroid 22-alpha-hydroxylase T9L24.44 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C/Accession: H96759

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A/Accession: H96759

A>Status: preliminary

A/Molecule type: DNA

A/Residues: 1-512 <STO>

A/Cross-references: GB:AE005173; NID:g11120803; PIDN:AAG30983.1; GSPDB:GN00141

C/Genetics:

A/Gene: T9L24.44

A/Map position: 1

Query Match 12.8%; Score 333.5; DB 2; Length 512;

Best Local Similarity 24.9%; Pred. No. 6.6e-18;

Matches 124; Conservative 84; Mismatches 201; Indels 89; Gaps 17;

QY 6 LLSALCTFVLPLLLFLAIAIKLMDLYCVSGDRSCALPLPGTMGPFPEFGET----- 57

Db 12 LSVSSSTTFLLAIIIFLLA-----GIARRKRRAPHRLLPGSGRWPLIGDTFAMLNVA 64

QY 58 -----LQVYLQRRKFLQMKRRKYGFYKTHLFGPRTVRVMGAD 95

Db 65 GSHPSFVEKQIKKFVSLICSVLLILKRPDNGSFNEIRYGRIFCSLFGKWAIVSADPD 124

QY 96 NVRIILGDDRLVSVHWPASVRTILSGCLSNLHDSHKOKKVIIMRAFSEALECY--- 152

Db 125 FNRFIQNEGKLFQSSYPKSFRLDVGKGVITVHGDOQRRLHSIASSMMRHQDKTHFLE 184

QY 153 -VPVITEEVGSSLEQWLSGGERGLVYPEVKRLMFIAMRILLCCEPQLAG-DGDSE-QQ 209

Db 185 VIPVVMQLTLSNFKD---GEVLL-----QDICKVAIHLMVN---QLLGVSSESEVDE 232

QY 210 LVEAFEEEMTRNLFSLPIDVPFSGLYRGMKARNLIHARIEQNIIRAKICGLRASE-AGGCK 268

Db 233 MSQLFSDPYDGLSVPIDLPGFYTNKAMKARKEIIRKINTIEKRLONKAASDTAGNG-- 290

QY 269 DALQLIETHSWGERGLDMQALKQSTTELLFGGHETTASATSLITYLGLYPHYLVOKVRE 328

Db 291 -VLGRLEE-----ESLPNESMADFIINLLFAGNETTSKTMFAVYFLTHCPKAMTQLE 344

QY 329 E--LKSGLCKSNQDNKLDMEILEQLYIGCVIKETLRLNPPVPG-----FRVALKTF 381

Db 345 EHDRLAGML--TWQDYK-----TMDFTQCVIDETLRL-----GGIAIWLREAKEDV 390

QY 382 ELNGYQIPKGMVYISICDTHDVAEIFTNKEEFNPDPSAPHPEDASRFS---FIPFGG 437

Db 391 SYQDYVIPKGCFFVPLSAVHLDESYKESISFNPMRWLDPETOQRNWRMTSPFYCPFGG 450

QY 438 GLRSCVKGKEFAKILLKIE 455

Db 451 GTRFCPGAELARLQIALF 468

RESULT 15

S41192

cytochrome P450 4D2 - fruit fly (Drosophila melanogaster)

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Drosophila melanogaster

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C/Accession: S41192; S34291

R/Frolov, M.V.; Alalortsev, V.E.

submitted to the EMBL Data Library, December 1993

A/Description: A cluster of cytochrome P450 genes in the X-chromosome of Drosophila m

A;Reference number: S41192
A;Accession: S41192
A;Molecule type: DNA
A;Residues: 1-496 <FRO>
A;Cross-references: EMBL:X75955; NID:g439650; PIDN:CAA53568.1; PID:g439651
A;Experimental source: strain Oregon R
R;Frolov, M.V.; Alatorsev, V.E.
submitted to the EMBL Data Library, June 1993
A;Description: Cluster of cytochrome P-450 genes on the X-chromosome in Drosophila melanogaster
A;Reference number: S34291
A;Accession: S34291
A;Molecule type: DNA
A;Residues: 'A',31-496 <FRW>
A;Cross-references: EMBL:Z23005; NID:g312903; PIDN:CAA80549.1; PID:g312904
A;Experimental source: strain Oregon R
C;Genetics:
A;Gene: Cyp4d2
A;Cross-references: FlyBase:FBgn0011576
A;Map position: X
A;Introns: 53/1; 182/1; 205/2; 392/1
C;Superfamily: human cytochrome P450 CYP4B1; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; oxidoreductase
F;308-471/Domain: cytochrome P450 homology <P45>
F;449/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match		12.6%	Score 328.5;	DB 1;	Length 496;
Best Local Similarity		24.7%	Pred. No. 1.5e-17;		
Matches 124;		Conservative 91;	Mismatches 232;	Indels 55;	Gaps 13;
QY	18 LLLFLAIAIKLMDLYCVSGRDRSCALPLPPTMGFPFEGETLQ-----VLQRRKFLQMKRR 73				
Db	9 LLVAFATLLLMDFLW-----RRRGNGILPGRPRLPFLGNLIMYRGLDPEQIMDFVKKNR 63				
QY	74 KYGFIYKTHLFGPRTVRVMGADNVRILLGDDRLVSVHWPASVRTILSGCLSNLHDSH 133				
Db	64 KYGRLYRVWILLQ LAVFSTDPDRIEFVLSQOHITKNLKLNCWLGDLIMSTGRKW 123				
QY	134 KQKKVIMRAFSREALECYVPVITEVGSLSLEWLS--CGERGLVYPEVKRLMERIAMR 191				
Db	124 -GRRKITTPTTFHFKILEQVEIFDQSAVWEQLOSRRDGMTPIINFVICLTALDITAE 182				
QY	192 ILIGCE-----POLAGDGSEQLVEAFEEEMTRN--LFSL--PIDVPFSGLYRCM 237				
Db	183 TAMGTKINAQKNPNLPYVQAVNDVTNIIKRFIHAMQRVDWIFRLTQPTAK-----RQD 237				
QY	238 KARNLIHARIEQNIRAKIGL-----RASEAGQCKDALQOLIIEHSWGERGLD 286				
Db	238 KAIKVMHDTTENIRERRETLVNNSKETTPREEVNFLLGQKRRMALDVLLOSTIDGAPLS 297				
QY	287 MQALKOSTELLFGGHETTASATSLTYLGLYPHVLQKVREELKSKGLLCKSNQDNKLD 346				
Db	298 DEDIREVDTFMEFGHDTTSAISFCLEYISRHPVEVQRLQOEIRD--VLGEDRKSPVT 354				
QY	347 MEILEQLKYIGCVIKETLRLNPPVPGGFYVALKTFELNGYQIPKGNVYISICDTHDVAE 406				
Db	355 LRDLGELKFMENVIKESLRLHPVPVPMIGWFAEDVEIRGKHIPAGTNFTMGIFVLLRDPE 414				
QY	407 IFTNKEEFNDPDRSAPHPEASRFSIFPGGGLRSCVSGKEFAKILKIFTVELARHCDWQ 466				
Db	415 YFESPDEFRRPERFDADVPQ-IHPYATIPFSAGPRNCIGOKFAMLEMKSTVSKLLRH--FE 471				
QY	467 LLNGPPTMKTS-----PTVY 481				
Db	472 LLPLGPEPRHSMNIVCGRPTAF 493				

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:39:39 ; Search time 62.8 Seconds
(without alignments)
271.098 Million cell updates/sec

Title: US-09-668-482-4
Perfect score: 2611
Sequence: 1 MGLPALASALCTFVLPILL.....PTVYPVDNLPARTFHGEI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 segs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2601	99.6	497	1	CP26_HUMAN	O43174 homo sapien
2	2437	93.3	497	1	CP26_MOUSE	O55127 mus musculu
3	1738	66.6	492	1	CP26_BRARE	P79739 brachydanio
4	463.5	17.8	490	1	CP83_ARATH	O23051 arabidopsis
5	444	17.0	464	1	CP85_LYCES	O43147 lycopersico
6	433.5	16.6	472	1	CP901_ARATH	O42569 arabidopsis
7	400	15.3	492	1	CP136_MYCTU	P95099 mycobacteri
8	388	14.9	519	1	CP881_MAIZE	O43246 zea mays (m
9	338.5	13.0	476	1	CP137_MYCTU	O69653 mycobacteri
10	331	12.7	501	1	CP42_DROME	O27589 drosophila
11	325.5	12.5	511	1	CP45_RABIT	P14579 oryctolagus
12	325	12.4	503	1	CP51_RAT	O64654 rattus norv
13	324.5	12.4	512	1	CP421_DROME	P33269 drosophila
14	322	12.3	520	1	CP84_ARATH	O42600 arabidopsis
15	321	12.3	503	1	CP51_PIG	O46420 sus scrofa
16	320.5	12.3	512	1	CP421_DROSI	O16805 drosophila
17	320	12.3	472	1	CP13B_MYCTU	O53765 mycobacteri
18	318.5	12.2	511	1	CP4C_BIADI	P29981 blaberus di
19	317	12.1	518	1	CP3R_ONCMY	O42563 oncorhynch
20	315.5	12.1	511	1	CP47_RABIT	P14581 oryctolagus
21	314.5	12.0	493	1	CP132_DROME	O9vgb4 drosophila
22	312.5	12.0	500	1	CP46_MOUSE	O9wvk8 mus musculu
23	312	12.0	519	1	CP4Y_HUMAN	O02928 homo sapien
24	312.5	11.9	503	1	CP51_HUMAN	O16850 homo sapien
25	311.5	11.9	505	1	CP48_DROME	O9vs79 drosophila
26	308	11.8	507	1	CP3S_BOVIN	P79102 bos taurus
27	305.5	11.7	453	1	CP51_WHEAT	P93596 triticum ae
28	305.5	11.7	500	1	CP46_HUMAN	O9y6a2 homo sapien
29	305.5	11.7	509	1	CP41_RAT	P08516 rattus norv
30	305.5	11.7	530	1	CP51_YEAST	P10614 saccharomyc
31	304.5	11.7	492	1	CP51_SORBI	P93846 sorghum bic
32	304.5	11.7	537	1	CP6_RAT	P51871 rattus norv
33	303.5	11.6	512	1	CP51_CUNEL	O9uvc3 cunninghame

34	303	11.6	441	1	CP138_MYCTU	P96813 mycobacteri
35	302.5	11.6	459	1	CPXN_ANASP	P29980 anabaena sp
36	301	11.5	496	1	CP4AE_DROME	O46054 drosophila
37	301	11.5	528	1	CP51_CANTR	P14263 candida tro
38	300	11.5	483	1	CP311_DROME	O9vyq7 drosophila
39	300	11.5	511	1	CP772_SOLME	P37124 solanum mel
40	297.5	11.4	510	1	CP46_RABIT	P14580 oryctolagus
41	296.5	11.4	494	1	CP134_DROME	O9vg40 drosophila
42	295	11.3	520	1	CPF3_HUMAN	O08477 homo sapien
43	294.5	11.3	487	1	CP135_DROME	O9vgb5 drosophila
44	294	11.3	506	1	CP44_RABIT	P10611 oryctolagus
45	292.5	11.2	510	1	CP421_DROME	O9w011 drosophila

ALIGNMENTS

RESULT 1
CP26_HUMAN
ID CP26_HUMAN STANDARD; PRT; 497 AA.
AC O43174;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RAI) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=97373542; PubMed=9228017;
RX White J.A., Beckett-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,
RA Jones G., Petkovich M.;
RT "CDNA cloning of human retinoic acid-metabolizing enzyme (hp450RAI)
RT identifies a novel family of cytochromes P450.";
RL J. Biol. Chem. 272:18538-18541(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98380037; PubMed=9716180;
RA Sonneveld E., van den Brink C.E., van der Leede B.M., Schulkes R.K.,
RA Petkovich M., van der Burg B., van der Saag P.T.;
RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for
RT all-trans-RA and can be induced through RA receptors in human breast
RT and colon carcinoma cells.";
RL Cell Growth Differ. 9:629-637(1998).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=99045433; PubMed=9826557;
RA Trofimova-Griffin M.E., Juchau M.R.;
RT "Expression of cytochrome P450RAI (CYP26) in human fetal hepatic and
RT cephalic tissues.";
RL Biochem. Biophys. Res. Commun. 252:487-491(1998).
CC -!- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEROISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN ADULT LIVER, HEART,
CC PITUITARY GLAND, ADRENAL GLAND, PLACENTA AND REGIONS OF THE BRAIN.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
DR EMBL: AF005418; AAB88881.1; -.
DR MIM; 602239; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56162 MW; EAB6B84B24B2EAB3 CRC64;

Query Match 99.6%; Score 2601; DB 1; Length 497;
Best Local Similarity 99.6%; Pred. No. 2e-171;
Matches 495; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPPLLEFLAAIKLMDLYCVSGRDRSCALPLPPTMGFPFFGETLQM 60
Db 1 MGLPALLASALCTFVLPPLLEFLAAIKLMDLYCVSGRDRSCALPLPPTMGFPFFGETLQM 60
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGRPTRVMGADNVRILLGDDRLLVSVHWPASVRIIL 120
Db 61 VLQRRKFLQMKRRKYGFIYKTHLFGRPTRVMGADNVRILLGDDRLLVSVHWPASVRIIL 120
QY 121 GSGCLSNLHDSHKQKKVIMRAFSREALCEYVPVITEEVGSSLEQWLSGGERGLVYPE 180
Db 121 GSGCLSNLHDSHKQKKVIMRAFSREALCEYVPVITEEVGSSLEQWLSGGERGLVYPE 180
QY 181 VKRLMFRIAMRILLGCEPQLAGDGDSEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKAR 240
Db 181 VKRLMFRIAMRILLGCEPQLAGDGDSEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKAR 240
QY 241 NLIHARIEQNIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKOSSTELLFG 300
Db 241 NLIHARIEQNIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKOSSTELLFG 300
QY 301 GHETTASAATSLITYLGLYPHVLOKVREELKSKGLLCKSNQDNKLDMEILEQLKYIGCVI 360
Db 301 GHETTASAATSLITYLGLYPHVLOKVREELKSKGLLCKSNQDNKLDMEILEQLKYIGCVI 360
QY 361 KETLRLNPVPVPGGFRVALKTFELNGYQIPKGWNVYISICDTHDVAEIFTNKEEFNDRFS 420
Db 361 KETLRLNPVPVPGGFRVALKTFELNGYQIPKGWNVYISICDTHDVAEIFTNKEEFNDRFM 420
QY 421 APHPEDASRSFIFPGGGLRSCVKGKEFAKILKIFVELARHCDWQLLNGPPTMKTSPTY 480
Db 421 LPHPEDASRSFIFPGGGLRSCVKGKEFAKILKIFVELARHCDWQLLNGPPTMKTSPTY 480
QY 481 YPVDNLPARFTFHGEI 497
Db 481 YPVDNLPARFTFHGEI 497

RESULT 2
CP26_MOUSE STANDARD; PRT; 497 AA.
AC 055127;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26 OR P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H;
RX MEDLINE=97392446; PubMed=9250660;
RA Fujii H., Sato T., Kaneko S., Gotoh O., Fujii-Kuriyama Y., Osawa K.,

RA Kato S., Hamada H.;
RT "Metabolic inactivation of retinoic acid by a novel P450
RT differentially expressed in developing mouse embryos.";
RL EMBO J. 16:4163-4173(1997).
RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=98113212; PubMed=9442090;
RA Abu-Abed S.S., Beckett B.R., Chiba H., Chithalen J.V., Jones G.,
RA Metzger D., Chambon P., Petkovich M.;
RT "Mouse P450RA1 (CYP26) expression and retinoic acid-inducible retinoic
RT acid metabolism in F9 cells are regulated by retinoic acid receptor
RT gamma and retinoid X receptor alpha.";
RL J. Biol. Chem. 273:2409-2415(1998).

CC -!- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -!- INDUCTION: BY RETINOIC ACIDS (RA).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL: Y12657; CAA73206.1; -.
DR MGD; MGI:1096359; Cyp26.
DR InterPro: IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56177 MW; 33B07D7C29134471 CRC64;

Query Match 93.3%; Score 2437; DB 1; Length 497;
Best Local Similarity 93.4%; Pred. No. 3.6e-160;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPPLLEFLAAIKLMDLYCVSGRDRSCALPLPPTMGFPFFGETLQM 60
Db 1 MGLPALLASALCTFVLPPLLEFLAAIKLMDLYCVSSRDRSCALPLPPTMGFPFFGETLQM 60
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGRPTRVMGADNVRILLGDDRLLVSVHWPASVRIIL 120
Db 61 VLQRRKFLQMKRRKYGFIYKTHLFGRPTRVMGADNVRILLGHRLLVSVHWPASVRIIL 120
QY 121 GSGCLSNLHDSHKQKKVIMRAFSREALCEYVPVITEEVGSSLEQWLSGGERGLVYPE 180
Db 121 GAGCLSNLHDSHKQKKVIMQAFSREALQCYLVIAEYVSSCLEQWLSGGERGLVYPE 180
QY 181 VKRLMFRIAMRILLGCEPQLAGDGDSEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRGVKAR 240
Db 181 VKRLMFRIAMRILLGCEPQAGGEDEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRGVKAR 240
QY 241 NLIHARIEQNIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKOSSTELLFG 300
Db 241 NLIHARIEQNIRAKIRLQATEPDGCKDALQLLIEHSWGERLDMQALKOSSTELLFG 300
QY 301 GHETTASAATSLITYLGLYPHVLOKVREELKSKGLLCKSNQDNKLDMEILEQLKYIGCVI 360
Db 301 GHETTASAATSLITYLGLYPHVLOKVREELKSKGLLCKSNQDNKLDMETLEQLKYIGCVI 360
QY 361 KETLRLNPVPVPGGFRVALKTFELNGYQIPKGWNVYISICDTHDVAEIFTNKEEFNDRFS 420
Db 361 KETLRLNPVPVPGGFRVALKTFELNGYQIPKGWNVYISICDTHDVAIDIFTNKEEFNDRFI 420


```

Best Local Similarity 27.6%; Pred. No. 8.5e-25;
Matches 132; Conservative 85; Mismatches 234; Indels 27; Gaps 11;

QY 25 IKLMDLYCVSGDRSCALDPBGTMGPFPEFGETLQWVLQRR-----KFLQMKRRKYG--F 77
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 29 VNWLWYESSLGENRHY---LPPGDLGWPFIGNMLSLRAFKTSDDPSFTRTLKRYGPKG 85
QY 78 IYKTHLFGRTVRYVMGADNVRRILLGDDRLLVSVHWMPASVRTILGSGCLSNLHDSHKQK 137
   ||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 86 IYKAHMFNGPSIIVTTSIDTCRRVLTDDAF-KPGWPTSTMELIGRKSFGVGISFEEHKRLR 144
QY 138 KVIMRAFS-REALCEYVPVITEEVGSSLEQWLSCGERGLLVYPEVKRLMFIAMRILLGC 196
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 145 RLTAAPVNGHEALSTYIPYIEENVITVLDKWTMGEEFELTH--LRKLTFRIMYIFLSS 202
QY 197 EPOLAGDGDSEQQLVEAFEMETRNLFSLPIDVPFSGLYGMRKARNLIHARIEONIRAKIC 256
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 203 ESENVMDA-----LERETALNYGVRAVAVNIPGFAYHRALKARKTLVAAF-QSIVTERR 256
QY 257 GLRASEAGQGCKDALQLLIEHSWGERLDMQALKQSTELLFGHETTASATSLITYL 316
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 257 NQKQNILSNKKMDLNLNVKDEDGKTLDDDEIIDVLLMYLNAGHSSGHTIMATVFL 316
QY 317 GLYPHVLQKVREELKSKGLCKSNQDNKLDMEILE--QLKYIGCVIKETLRLNPPVPGGF 374
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 317 QEHPEVLQRAKAE---QEMILKSREPEGKGLSLKETRKMEELSQVDETLRVITFSLTAF 373
QY 375 RVALKTFELNGYQIPKGMNVYISICDTHDVAEIFTNKEEFNDRFSAPHDPEDASRESFIP 434
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 374 REAKTDEVEMNGYLIPKGMKVLTWFRDVHIDPEVFPPDRKEDPARWDNGFVPKAG--AFLP 431
QY 435 FGGGLRSCVCGKEFAKILKIFTVELARHCDWQLNGPPMTKTSPTVYPVDNLPARFTH 492
   || | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 432 FGAGSHLCPGNDLAKLETISIFLHFLKLYQVKRSNPCEPYMYLPHTRPTDNCLEARISY 489

RESULT 5
CP85_LYCES
ID CP85_LYCES STANDARD; PRT; 464 AA.
AC Q43147;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 85 (EC 1.14.--.) (DWARF PROTEIN).
GN CYP85 OR D.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. GCR758;
RX MEDLINE=96266705; PubMed=8672892;
RA Bishop G.J., Harrison K., Jones J.J.G.D.;
RT "The tomato Dwarf gene isolated by heterologous transposon tagging
   encodes the first member of a new cytochrome P450 family.";
RL Plant Cell 8:959-969(1996).
CC -|- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation--
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U54770; AAB17070.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.

```

```

KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BINDING 414 414 HEME (BY SIMILARITY).
SQ SEQUENCE 464 AA; 53706 MW; D2B21AAB7B14E94 CRC64;

Query Match 17.0%; Score 444; DB 1; Length 464;
Best Local Similarity 27.4%; Pred. No. 1.7e-23;
Matches 128; Conservative 94; Mismatches 203; Indels 42; Gaps 12;

OY 18 LLLFLAIAKLMDLYCVSGDRSCALPLPGTGMFPFEGETLQAVLQRRKFLQMKRRKYGF 77
   | : | : | : : : | | | | | | | : : : | : |
DB 14 LCIFCTALLRWNOVKYNQKN-----LPPGTMGWPLFGETTEFLKLGPSFMKNQARAYGS 67

OY 78 IYKTHLFGRPTRVMGADNVRILLGDDRLVSVHWPASVRTILIGSGCLSNLHDSHKQRK 137
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 68 FFKSHILGCPPTIVSMDSELNRYILLVNEAKGLVPGYPQSMIDLIGKCNIAVNGSAHKYMR 127

OY 138 KVIMRAFS---REALCEYVPVITEEVGSSLEQWLSCGERGLVYPEVKRLMFRIAMRIL 193
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 128 GALLSLISPTMIRDQL--LPKIDEFMRSHLTNW--DNKVIDIQEKTNKMFAFLSLKQI 181

OY 194 LGCE-PQLAGDGDSEQLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKAR---NLIHARIE 248
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 182 AGIESTSLA-----QEFMSEEFNLVLGTLSPINLPNTNHRGFQARKIIVNLRRLIE 235

OY 249 QNIRAKICGLRASEAGQCKDALQLLIEHSWGERGLDMQALKOSSTELLFGHETTASA 308
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 236 ER-----RASKETQ--HDMGLYLMEAEATRFKLTDEMDILIT-ILYSGYETVST 284

OY 309 ATSLITYLGLYPHYLQKVRRELKSKGLCKSNQDNKLDMEILEQLKYGIVIKETRLNP 368
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 285 SMAAVKYLHDHPKVLLELRKE--HMAIREKKKPEDPIDYNDYRSMRFTRAVILETSRLAT 342

OY 369 PVPGEFRVALKTFELNGYQIPKGNVYISICDTHDVAEIFTNKEFNPDREFSAPHPEDAS 428
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 343 IVNGVLRRKTTQDMEINGYIIPKGWRIYVYTRRELNDPRLYPDPYSFNPWRMDKSLHEQN 402

OY 429 RFSFIPEGGLRSCVSGKEFAKILLKIFTVELARHCDWQLNGPPTMK 475
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 403 --SFLVFGGTROCPGKRGELGVAEISTFLHYFVTKYRWEEIGDGDKLMK 447

RESULT 6
C901_ARATH STANDARD; PRT; 472 AA.
AC Q42569;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 90A1 (EC 1.14.--.).
GN CYP90A1 OR CYP90 OR CPD.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Virdidiplantae; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=96200769; PubMed=8612270;
RA Szekeres M., Nemeth K., Koncz-Kalman Z., Mathur J., Kauschmann A.,
RA Altman T., Redei G.P., Nagy F., Schell J., Koncz C.;
RT "Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
RT controlling cell elongation and de-etiolation in Arabidopsis.";
RL Cell 85:171-182(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to license@sib.ch.

CC EMBL; X87367; CAA60793.1; -
DR EMBL; X87368; CAA60794.1; -
DR InterPro; IPR001128; -
DR Pfam; PF00067; P450; 2.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Heme.
FT BINDING 418 418 HEME (BY SIMILARITY).
SQ SEQUENCE 472 AA; 53785 MW; 41A73F46D64E343F CRC64;

Query Match 16.6%; Score 433.5; DB 1; Length 472;
Best Local Similarity 26.5%; Pred. No. 9.3e-23;
Matches 129; Conservative 90; Mismatches 228; Indels 39; Gaps 13;

QY 19 LFLFAIKLMDLYCVSGRDRSCALPLPGTMGFPFGETLQMV-----LQRRKFLQMKRR 73
DB 7 LLLLSIAAGFLLLLR-RTRRRMGLPPGSLGLPLIGETPQLIGAYKTENPEPFIDERVA 65
QY 74 KYGFIYKTHLFGRTVVRVMGADNVRRIILGDDRLVSVMHPASVRTLIGSGCLSNLHDSH 133
DB 66 RYGSVFETHLFGEPITFSADPETNRFLQNEGKLFECSPASICNLGKSHLLMKGSLH 125
QY 134 KQKKVIMRAFSREALECYVPV-ITEEVGSSLEQWLSCGERGLLVPEVKRLMFRIAMRI 192
DB 126 KRMHSLTMSFANSSI IKDHLMDIDLRLVFNLDSSWS-----RVLLMEAKKITFELTVKQ 181
QY 193 LIGCEPOLAGDSEQQLVEAFEEEMTRNLFSLPIDVPESGLYR-GMKARNLIHARIEQNI 251
DB 182 LMSFDP-----GEWSESLRKEYLLVIEGFSLPLPL-FSTTYRKAIQARR---KYAEAL 231
QY 252 RAKICGLR--ASEAGGCKDALQLI--EHSWGERLDMQALKOSSTELLFGGHETTAS 307
DB 232 TVVVMKRREEEEGAERKKMDMLAALDAGFSDEIYDF-----LVALLVAGYETTST 285
QY 308 AATSLITFYLGLYPHVLQKVRRELKSKGLCKSNQDNKLDMEILEQLKYIGCVIKETRLN 367
DB 286 IMTLAVKFLTETPLALQAKKE--HEKIRAMKSDSYSLEWSDYKSMPTQCVVNETLRYA 343
QY 368 PVPVGGFRVALKTFELNGYQIPKGMNVYISICDTHDVAEIFTNKEEFNPDRFSAPHPEDA 427
DB 344 NIIGVFERRAMTDEVIKGYKIPKGMVFSSFRVAVHLDPNHFKDARTENPMRWQSNSTVTG 403
QY 428 SRFSFIPFGGLRSCVCKEFAKILKIFTVELARHCDWQLNGRPPTMTKTSPTVYVPDNL 487
DB 404 PSNVFTFPGGGPRLCPGYELARVALSVFLHRLVTGFSW-----VPAEQDKLVFFPTTRTQ 458
QY 488 ARFTHE 493
DB 459 KRYPIF 464

RESULT 7
C136_MYCTU STANDARD; PRT; 492 AA.
ID C136_MYCTU
AC P95099;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PUTATIVE CYTOCHROME P450 136 (EC 1.14.-.-).
GN CYP136 OR RV3059 OR MTCY22D7.22C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).

CC EMBL; Z83866; CAB06263.1; -
DR TubercuList; RV3059; -
DR InterPro; IPR001128; -
DR InterPro; IPR002397; -
DR InterPro; IPR002401; -
DR InterPro; IPR002403; -
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PRINTS; PR00385; P450.
DR PRINTS; PR00463; EP450I.
DR PRINTS; PR00465; EP450IV.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
FT BINDING 439 439 HEME (BY SIMILARITY).
SQ SEQUENCE 492 AA; 56227 MW; B0A78FCE95622F3D CRC64;

Query Match 15.3%; Score 400; DB 1; Length 492;
Best Local Similarity 28.2%; Pred. No. 1.9e-20;
Matches 137; Conservative 81; Mismatches 192; Indels 76; Gaps 21;

QY 37 DRSCALPLPPT-----MGFPFGETLQVNLQRRKFLQMKRRKYGYIKTH---LFG 85
DB 42 EKKLAEP-PPGSGLKPPVGDAGLPILGHMIEMLRGCPDYLM-----FLYKTKGPVVEG 93
QY 86 ----RPTVRVMGADNVRRIILGDDRLVSVH-WPASVRTLIGSGCLSNLHDSHKQKKVI 140
DB 94 DSAVLLPGVAALGPDAQYIYSNRNKDYSQQGWVPYIGPFFHRG-LMLDFEEHMFHRRIM 152
QY 141 MRAFSREALECYVPVITEEVGSSL-EQWLSCGERGLLVPEVKRLMFRIAMRIILGCEPQ 199
DB 153 QEAFVRSRLAGYLEQMDRVYSRVADWV-VNDARFLVYPAMKALTIDIASVMFGHEP- 210
QY 200 LAGDGDSEQQLVEAFEEEMTRN-----LFSLPIDVPESGLYRGMKARNLIH---ARIEQN 250
DB 211 -GTDHELVTKNKAFITITRAGNAVIRTSVP---PFT-WWRGLRARELLENYFTARVKE- 264
QY 251 IRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKOSSTELLFGGHETTASAAT 310
DB 265 -----RREASG---NDLLTYLCQTEDDDGNRFSADADIVNHMIFLMAAHDTSTSTAT 313
QY 311 SLITYLGLYPHVLQKVRRELKSKGLCKSNQDNKLDMEILEQLKYIGCVIKETRLNRPV 370
DB 314 TMAYQLAAHPWEQQRCDRESDRHG-----DGPLDIESLEQLSLDVNMESIRLVTPV 366
QY 371 PGCFRVALKTFELNGYQIPKGMNVYISICDTHDVAEIFTNKEEFNPDRFSAPHPED-DA 429
DB 367 QWAMRQTVRDETLGYYLPKGTNVIAYPGMNHRLEPMTDPLTFDPERFTEPRNEHKRHR 426
QY 430 FSFIPFGGLRSCVCKEFAK-----LKIFTVELARHCDWQLNGRPPTMTKTSPTVY 481
DB 427 YAFTPFGGCVHKCIGMVFQLEIKITLHRLRLRYRLDSRP-DYQ-----PRMDYSAMPI 480
QY 482 PVDNLP 487
DB 482 PVDNLP 487

```

Db      481 PMDGM 486

RESULT      8
C881_MAIZE

ID      C881_MAIZE      STANDARD;      PRT;      519 AA.
AC      Q43246;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      CYTOCHROME P450 88A1 (EC 1.14.-.-) (DWARF3 PROTEIN).
GN      CYP88A1 OR D3.
OS      Zea mays (Maize).
OC      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC      Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC      Andropogoneae; Zea.
OX      NCBI_TaxID=4577;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV. B73;
RX      MEDLINE=96004534; Pubmed=7549486;
RA      Winkler R.G., Helentjaris T.;
RT      "The maize Dwarf3 gene encodes a cytochrome P450-mediated early step
RT      in Gibberellin biosynthesis.";
RL      Plant Cell 7:1307-1317(1995).
CC      -1- PATHWAY: EARLY STEP IN GIBBERELLIN BIOSYNTHESIS.
CC      -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, DEVELOPING LEAVES, THE
CC      VEGETATIVE MERISTEM, AND SUSPENSION CULTURE CELLS.
CC      -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U32579; AAC49067.1; -.
DR      InterPro; IPR001128; -.
DR      Pfam; PF00067; P450; 1.
DR      PROSITE; PS00086; CYTOCHROME_P450; 1.
KW      Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT      TRANSMEM 1 21 POTENTIAL.
FT      BINDING 466 466 HEME (BY SIMILARITY).
SQ      SEQUENCE 519 AA; 57906 MW; 0F8977A024316D95 CRC64;

Query Match      14.9%; Score 388; DB 1; Length 519;
Best Local Similarity 27.2%; Pred. No. 1.4e-19;
Matches 126; Conservative 86; Mismatches 216; Indels 36; Gaps 14;

QY      44 LPPGTMGPFFFGFTLQWVLRK-----FLQMKRRKYG--FIYKTHLFGRTVRVMGADN 96
DB      72 LPPGEMGWPLVGGMAFLRAFKSGKPDAFIASFVRRRGRTGVYRSFMFSSPTVLVTTAEG 131
QY      97 VRRILLGDDRIVSVHMPASVRTILGSGCLSNLHDSHKQKKVIMRAFSR-EALECIYVP 155
DB      132 CKQVLMDDDAFVT-GWPKATVALVGPRSFVAMPYDEHRRIRKLTAAPIINGFDALTGYLPF 190
QY      156 ITTEVGSSLEQWLTSCGERGLLV-PEVKRLMFRIAMRTILGCEPQLAGDGDSEQOLVEAF 214
DB      191 IDRTVTSSLRAWADHG--GSVEFTELRLRMTFKIIYQIFLG----GADQATTRALERSY 243
QY      215 EEMTRNLFSLPIDVPESGLYRG-MKARNLIHARIEQNIRAKTCGLRASEAGGCKDALQL 273
DB      244 TELNYGMRAMAINLP-GFAYRGALRARRLVA-VLOGVLDERRAARAKGVSGGVDMMDR 301
QY      274 LIHSWGERGLDMQALKQSSTELLFGGHETTASATSLITYLGLYPHVLQKVRELKSK 333
DB      302 LIEAODERGRHLDDDEIIDLVMYMLNAGHSSGHTIMWATVFLQENPDMFARAKAEQEI 361
QY      334 GLICKSNQDNKLMDEITLQIKYIGCVIKETLRINPVPYGGFRVALKTFELNGYQIPGWN 393

```

```

Db      362  MRSIPSSQ-RGLTLBDFRKM EYLSQVIDELRLVNISFVSFRQATRVFVNGYLIPKGMK 420
QY      394  VIYSICDTHDVAELFTNKEEFNPDREFSAPHEDASRFSFIPIFGGLRSCVCGKEFAKILK 453
Db      421  VQLMYRSVHMDPOVYPDPTRKFDPSRWECHSPRAG---TFPLAFGLCARLCPGNDLAKLEIS 477
QY      454  IFT-----VELARHCDWQLLNGPPTMTKTSPTVYPVDNLPAFRT 491
Db      478  VFLHFLGLGYKLAR-----TNPRCVRVRYLPHRPVNDCLAKIT 515

RESULT  9
C137_MYCTU
ID      C137_MYCTU      STANDARD;      PRT;      476 AA.
AC      069653;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      PUTATIVE CYTOCHROME P450 137 (EC 1.14.-.-).
GN      CYP137 OR RV3685C OR MTW025.033C.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=H37RV;
RX      MEDLINE=98295987; PubMed=9634230;
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA      Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA      Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA      Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA      Taylor K., Whitehead S., Barrell B.G.;
RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence.";
RL      Nature 393:537-544(1998).
CC      -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AL022121; CA18007.1; -.
DR      Tuberculist; RV3685C; -.
DR      InterPro; IPR001128; -.
DR      InterPro; IPR002397; -.
DR      InterPro; IPR002401; -.
DR      Pfam; PF00067; P450; 1.
DR      PRINTS; PR00359; BP450.
DR      PRINTS; PR00385; P450.
DR      PRINTS; PR00463; EP450I.
DR      PROSITE; PS00086; CYTOCHROME_P450; 1.
KW      Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
FT      BINDING 422 422 HEME (BY SIMILARITY).
FT      SEQUENCE 476 AA; 52265 MW; 4FF0DBC89D8D0548 CRC64;

Query Match      13.0%; Score 338.5; DB 1; Length 476;
Best Local Similarity 25.2%; Pred. No. 3.1e-16;
Matches 102; Conservative 76; Mismatches 178; Indels 49; Gaps 10;

QY      65  RKFLQMKRRKYGFIYKTHLFGRPTRVVMGADNVRR-----ILGDDRLVSVHWPASV 116
Db      68  RRMLRLIRR-YGPIMTMPILSLGDVAIVSDSALAKEVFAPTVDVLLGEGV----GPAA- 121

```


DR FlyBase; FBgn0011576; Cyp4d2.
DR InterPro; IPR001128; -.
DR InterPro; IPR002402; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PRINTS; PR00464; EP45011.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum; Polymorphism.
FT BINDING 449 449 HEME (BY SIMILARITY).
FT VARIANT 163 163 K -> M (IN STRAINS CAM-44, CAM-48 AND
FT CONFLICT 2 25 BERKELEY).
FT CONFLICT 30 30 MISSING (IN REF. 2).
FT CONFLICT 160 160 I -> A (IN CAA80549).
FT CONFLICT 487 501 A -> R (IN REF. 1).
FT CONFLICT 501 501 LRSANGVHLGKPR -> CGRPTAFILA (IN REF.
SQ SEQUENCE 501 AA; 57914 MW; 8F8D98DBF39FD1A2 CRC64;

Query Match 12.7%; Score 331; DB 1; Length 501;
Best Local Similarity 24.7%; Pred. No. 1.1e-15;
Matches 120; Conservative 89; Mismatches 231; Indels 46; Gaps 11;

QY 18 LLEFLAIIKLMDLYCVSGDRSCALPLPPGTMGFFGETLQM---VLQRRKFLQMKRR 73
DB 9 LIVAFAFLLIMDFLW-----RRRGNGILPGRPLPFLGCLNLMYRGLDPEQIMDFVKKQR 63
QY 74 KYGEIYKTHLFGRPVRYMGADNVRRIILGDDRIVSVHWPASVRTILGSGCLSNLHDSH 133
DB 64 KYGRLYRVWLHQLAVFSDPRDIEFVLSQOHITKNNLYKLLNCWLGDLGSLMSTGRKH 123
QY 134 KQKKVIMRAFSREALECYVPVITEEVGSSLEQWLS--CGERGLLVPEVKRLMERIAMR 191
DB 124 -GRKIITPTFHKILEQVEIIFDQSAVMEQLQSRADGKTPINIFVICLTALDITAE 182
QY 192 ILGCE-----POLAGDGDSEQLVEAFEEEMTRN---LFSL--PIDVPFSGLYRGM 237
DB 183 TAMGTKINAQKNPNLPYQAVNDVTNLIKRFIHAMQRYDWIFRLTQPTAK-----RQD 237
QY 238 KARNLIHARIEQNIRAKICGL-----RASEAGOCCKDALQLLIEHSWGERLD 286
DB 238 KAIVKMHDFTEITIRERRETLVNSKETTPPEEVNPLGQKRMAALLDVLLOSTIDGAPLS 297
QY 287 MQALKQSTELLEFGHETTASATSLITYLGLYPHYLQVREELKSKGLCKSNQDNKLD 346
DB 298 DEDIREVDTEMEFGHDTTSAISFCLYEISRHPVEQORLQOEIRD--VLGEDRKSPVT 354
QY 347 MEILEQLKYIGCVIKETLRNPPVPGGFVALKTFEINGYQIPKGMNVIYSICDTHDYAE 406
DB 355 LRDLGELKFMENVIKESIRHPVPMIGRWFAEDVEIRGKHIPAGTNFTMGIFVLRDPE 414
QY 407 IFTNKEEFNDRFSAPHEDASRESFIPEFGGLRSCVSGKEFAKILLKFTVELARHCDWQ 466
DB 415 YFESPDDEFRPERFDADVQ-IHPYAIIPFSAGRNCIGQKFAMLEMKSTVSKLRRHELL 473
QY 467 LLNGPP 472
DB 474 PLGPER 479

RESULT 11
CP45_RABIT
ID CP45_RABIT STANDARD; PRT; 511 AA.
AC P14579;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CYTOCHROME P450 4A5 PRECURSOR (EC 1.14.15.3) (CYP1A5) (LAURIC ACID
DE OMEGA-HYDROXYLASE).
GN CYP4A5.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=90254128; PubMed=2340280;
RA Johnson E.F., Walker D.L., Griffin K.J., Clark J.E., Okita R.T.,
RA Meurhoff A.S., Masters B.S.;
RT "Cloning and expression of three rabbit kidney cDNAs encoding lauric
RT acid omega-hydroxylases.";
RL Biochemistry 29:873-879(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=91192021; PubMed=2013275;
RA Yokotani N., Kusunose E., Sogawa K., Kawashima H., Kinasaki M.,
RA Kusunose M., Fujii-Kuriyama Y.;
RT "cDNA cloning and expression of the mRNA for cytochrome P-450kd which
RT shows a fatty acid omega-hydroxylating activity.";
RL Eur. J. Biochem. 196:531-536(1991).
CC -I- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
CC ACIDS, AND XENOBIOTICS.
CC -I- CATALYTIC ACTIVITY: OCTANE + REDUCED RUBREDOXIN + O(2) = 1-OCTANOL
CC + OXIDIZED RUBREDOXIN + H(2)O.
CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -I- INDUCTION: P450 CAN BE INDUCED TO HIGH LEVELS IN LIVER AND OTHER
CC TISSUES BY VARIOUS FOREIGN COMPOUNDS, INCLUDING DRUGS, PESTICIDES,
CC AND CARCINOGENS.
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M28655; AAA31229.1; -.
CC EMBL; X57209; CAA40493.1; -.
CC PIR; A34260; A34260.
CC InterPro; IPR001128; -.
CC InterPro; IPR002402; -.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PRINTS; PR00464; EP45011.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT PROPEP 1 4 PROBABLE.
FT CHAIN 5 511 CYTOCHROME P450 4A5.
FT BINDING 458 458 HEME (BY SIMILARITY).
FT CONFLICT 435 435 G -> S (IN REF. 2).
FT CONFLICT 477 477 V -> L (IN REF. 2).
SQ SEQUENCE 511 AA; 58357 MW; 11D174BFC8BFA268 CRC64;

Query Match 12.5%; Score 325.5; DB 1; Length 511;
Best Local Similarity 26.2%; Pred. No. 2.6e-15;
Matches 137; Conservative 79; Mismatches 221; Indels 85; Gaps 21;

QY 3 LPALLASALCTFVL--PLILFLAIIKL-----WDLYCVSGDRSCALPLPPGTMGFFFG 55
DB 11 LPGLISGLLQVAAALLGLLLILKAQOLYLRQWLRLAQ-----QFPCCP----- 55
QY 56 ETIOWVLQRRKFLQMKRRKYGFYKTHLFGRPVRYMGADNVR-----RILGDDR 106
DB 56 -FHWLLGHSREFQMNQELQOLIKWVEKEPRACPHWIGGNKVRVQLYDPDYMKVILGRSD 113
QY 107 LVSVHWPASVRTILGSGCLSNLHDSHKKQKKVIMRAFSREALECYVPVITEEVGSSLEQ 166

Db 114 PKSRGTYTFVAPWIGYGLLL-LNGQPWEQHRRMLTPAFHYDILKPYIGLWVDSVOIMLDK 172
QY 167 W-LSCGERGLLVPEVKRLMFRIAMRILLCGPOLAGDGDSEQOLVEAFEE----- 217
Db 173 WEQLVSDSSLEVFQDISLMTLDTMKCAFSYQGSVOLDSRNSQSYIQAVGDLNMLVFAR 232
QY 218 TRNLESLPIDVPFSGLYR-----GKARNLIHARIEQONIRAKICGLRASEAGQGCK-- 268
Db 233 VRNIFH-----QSDTYRLSPESGRLSHRACQLAHEHTDRVIOQRKAQLQ--QEGELEKVR 285
QY 269 -----DALQLLIEHSMERGERLDMQALKQSSTELLFGHETTASATSLITYLGLYPHVL 323
Db 286 RKRRLDFLDVLLFAKMENGSSLSDDLRAEVDTFMEFGHDTASGVSMTFYALATHPEHQ 345
QY 324 QKVREELKSKGLCKSNODNKLDMELIQLKICVYIKETLRLNPPVPGGFR--VALKTF 381
Db 346 HRCREET--QGILL--GDGASITWEHLDMQMPYTTWCIKEAMRLYPVPVPAISRDLSSPVT 400
QY 382 ELNGYQIPKGNVYISICDTHDVAEIFTNKEFNPDRE--SAPHPEDASRFSIFPGGG 438
Db 401 P-DGRSLPKGFTVTLISYIGLHNPNVWPNPPEVDPGRFTPGSARHSH-----AFLPESGG 454
QY 439 LRSCVGKEFAKILKLTVELARHCDWQLNGBPPTMKTSP 479
Db 455 ARNCIGKQFAMNELKVAVALTLVR--FELL-PDPTRIKPT 492

RESULT 12

CP51_RAT STANDARD; PRT; 503 AA.

AC Q64654; Q64549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450 51 (EC 1.14.14.1) (CYP11) (P450L1) (STEROL 14-ALPHA DEMETHYLASE) (LANOSTEROL 14-ALPHA DEMETHYLASE) (LDM) (P450-14DM).
GN CYP51.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94296405; PubMed=8024575;
RA Aoyama Y., Funae Y., Noshiro M., Horiuchi T., Yoshida Y.;
RT "Occurrence of a P450 showing high homology to yeast lanosterol 14-demethylase (P450(14DM)) in the rat liver.";
RL Biochem. Biophys. Res. Commun. 201:1320-1326(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=98158318; PubMed=9498553;
RA Noshiro M., Aoyama Y., Kawamoto T., Gotoh O., Horiuchi T., Yoshida Y.;
RT "Structural and evolutionary studies on sterol 14-demethylase P450 (CYP51), the most conserved P450 monooxygenase: I. Structural analyses of the gene and multiple sizes of mRNA.";
RL J. Biochem. 122:1114-1121(1997).
RN [3]
RP SEQUENCE OF 18-503 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95394364; PubMed=7665087;
RA Sloane D.L., So O.Y., Leung R., Scarafia L.E., Saldou N., Jarnagin K., Swinney D.C.;
RT "Cloning and functional expression of the cDNA encoding rat lanosterol 14-alpha demethylase.";
RL Gene 161:243-248(1995).
CC -i- FUNCTION: CATALYSES C14-DEMETHYLATION OF LANOSTEROL; IT TRANSFORMS LANOSTEROL INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL.
CC -i- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH + OXIDIZED FLAVOPROTEIN + H(2)O.
CC -i- PATHWAY: CHOLESTEROL BIOSYNTHESIS.

CC -i- SUBCELLULAR LOCATION: MICROSOMAL (POTENTIAL).
CC -i- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D55681; BAA09529.1; -.
DR EMBL; AB004096; BAA20354.1; -.
DR EMBL; AB004087; BAA20354.1; JOINED.
DR EMBL; AB004088; BAA20354.1; JOINED.
DR EMBL; AB004089; BAA20354.1; JOINED.
DR EMBL; AB004090; BAA20354.1; JOINED.
DR EMBL; AB004091; BAA20354.1; JOINED.
DR EMBL; AB004092; BAA20354.1; JOINED.
DR EMBL; AB004093; BAA20354.1; JOINED.
DR EMBL; AB004094; BAA20354.1; JOINED.
DR EMBL; AB004095; BAA20354.1; JOINED.
DR EMBL; U17697; AAA87074.1; -.
DR InterPro; IPR001128; -.
DR InterPro; IPR002403; -.
DR Pfam; PF00067; P450; 2.
DR PRINTS; PR00465; EP450IV.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Transmembrane;
KW Heme; Cholesterol biosynthesis.
FT TRANSMEM 24 44 POTENTIAL.
FT BINDING 449 449 HEME (BY SIMILARITY).
FT CONFLICT 181 181 E -> K (IN REF. 3).
SQ SEQUENCE 503 AA; 56706 MW; 33D8F345FEE9CF21 CRC64;

Query Match 12.4%; Score 325; DB 1; Length 503;
Best Local Similarity 22.9%; Pred. No. 2.8e-15;
Matches 117; Conservative 95; Mismatches 236; Indels 62; Gaps 14;

QY 6 LLASAL--CTFVPLLL-LFLAIIKLMIDYCVSGDRSCALPLPGTMG-----FPPFG 55
Db 26 LLSTLLIACAFSLVYLFRLAV-----GHMVQLPAGAKSPPIYSPIFLG 72
QY 56 ETLQMLQRRKFLQMKRRKYGFITYKTHLFGRTVRVMGADNVRIIL--LGDDRIVSVHMP 113
Db 73 HAIAFGKSPTEFLNAYEKYGPVFSFTMGKTFITYLLGSDAALLFNSKNEDLNAEEVYG 132
QY 114 ASVRTILGGCLSNLHDSHKQRKKVIMAFSREALECYVPVYTEVYSSLEQWLSCGER 173
Db 133 RLTPVFGKGVADVPAVFLQKKILKSGLNIAHFQYVSIIEKAKEYFKSWGESGER 192
QY 174 GLVYPEVKRLMFRIAMRILLCGPOLAGDGDSEQOLVEAFEEEMRNLL--FS-----LP 225
Db 193 N-VFEALSELILTASHCLHGKE-----IRSQLNEKVAQLYADLDGFSHAAMLIP 242
QY 226 IDVPFSGLYRGMA---RNLIHARIEQONIRAKICGLRASEAGQCKDALQLLIEHSWER 281
Db 243 GWLPLPSFRRRDRAHREIKNIFYKAIQKRRLSK-----EPAEDILQTLIDSTYKD 292
QY 282 GERLDMQALKQSSTELLFGHETTASATSLITYLGLYPHVLQKYVREELKSKGLCKSNQ 341
Db 293 GRPLTDEIAGMLIGLLAGHTSSTTSAMWGFELARDKPLQDKCYLEQKT--VCGEDL 349
QY 342 DNKLDMELIQLKICVYIKETLRLNPPVPGGFRVALKTFELNGYQIPKGNVYISICDT 401
Db 350 P-PLTYEQKDLNLDRCIKETLRLRPPIMTMMMAKTPQTVAGYTIIPGHQCVSPTVN 408
QY 402 HDVAEIFTNKEFNPDREFSAPHPEDASRFSIFPGGGRLSCVSGKEFAKILKLTVELAR 461
Db 409 QRLKDSWERLDENPDRLQDNPASGEKFAVYFPGAGRHRICIGENFAVYQIKTIWSTMAR 468
QY 462 HCDWQLNGP-PTMKTSPTYVPVDNLPARF 490

Db 469 LYEFDLINGFSPSVNTTMIHTPENPVIRY 498

RESULT 13

C4D1_DROME STANDARD; PRT: 512 AA.

AC P33269; 018664; 018644; 018653; Q9W515; Q9W516;

DT 01-FEB-1994 (Rel. 28, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE CYTOCHROME P450 4D1 (EC 1.14.-.-) (CYP11D1).

GN CYP4D1 OR CYT-P450-D1 OR EG:87B1.1 OR CG3656.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephyroidae; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OREGON-R; TISSUE=Embryo;

RX MEDLINE=92297166; PubMed=1605861;

RA Gandhi R., Varak E., Goldberg M.L.;

RT "Molecular analysis of a cytochrome P450 gene of family 4 on the

RL DNA Cell Biol. 11:397-404(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=VARIOUS STRAINS;

RT "Evidence for non-neutral evolution around the cytochrome P450 gene

RT cluster on the Drosophila melanogaster X chromosome.";

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=OREGON-R;

RX MEDLINE=20196011; PubMed=10731137;

RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,

RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,

RA Dreano S., Gloux S., Ielaure V., Mottier S., Galibert F., Borkova D.,

RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,

RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,

RA Modolell J., Peter A., Schottler P., Werner M., Mourkioti F.,

RA Beinert N., Dowe G., Schafer U., Jackle H., Bucheton A.,

RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,

RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunder R.D.,

RA Glover D.M.;

RT "From sequence to chromosome: the tip of the X chromosome of D.

RT melanogaster.";

RL Science 287:2220-2222(2000).

RN [4]

RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Doherty Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

CC -1- FUNCTION: INVOLVED IN THE METABOLISM OF INSECT HORMONES AND IN THE

CC BREAKDOWN OF SYNTHETIC INSECTICIDES.

CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +

CC OXIDIZED FLAVOPROTEIN + H(2)O.

CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A

CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT, WITH THE

CC HIGHEST LEVELS OCCURRING DURING LATE LARVAL STAGES, THEN FALLING

CC DRASTICALLY DURING PUPARIATION.

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X67645; CAA47887.1; -

DR EMBL; Z98269; CAB10972.1; -

DR EMBL; AF016992; AAB71155.1; -

DR EMBL; AF016993; AAB71156.1; -

DR EMBL; AF016994; AAB71157.1; -

DR EMBL; AF016995; AAB71158.1; -

DR EMBL; AF016996; AAB71159.1; -

DR EMBL; AF016997; AAB71160.1; -

DR EMBL; AF016998; AAB71161.1; -

DR EMBL; AF016999; AAB71162.1; -

DR EMBL; AF017000; AAB71163.1; -

DR EMBL; AF017001; AAB71164.1; -

DR EMBL; AF017002; AAB71165.1; -

DR EMBL; AF017003; AAB71166.1; -

DR EMBL; AF017004; AAB71167.1; -

DR EMBL; AE003423; AAF45736.1; -

DR EMBL; AE003423; AAF45737.1; -

DR PIR; S25707; S25707.

DR FlyBase; FBgn0005670; Cyp4d1.

DR InterPro; IPR001128; -

DR InterPro; IPR002402; -

DR Pfam; PF00067; P450; 1.

DR PRINTS; PR00385; P450.

DR PRINTS; PR00464; EP450II.

DR PROSITE; PS00086; CYTOCHROME_P450; 1.

KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;

KW Endoplasmic reticulum; Alternative splicing; Polymorphism.

FT BINDING 456 456

FT VARSPLIC 1 186

FT

FT LVGHGHFIQKPPHEMVKRTFEFMEYTSKDVLYKWLGPPL

FT NVLMGNPKDVEVVLGTLRENDKAGEYKALEPWKEGLVSR

FT GRKWHKRRKIITPAHFETIDQFEVEFEKSRDLRNMED

FT RLKHGESGFSLYDWINICTMDT -> MWLLSTVLLAIIA

FT LEMRRFLRMRTIPGLPLPLGNNAHIFGLTPAEACIKIG


```
FT ELAERHGDTEGLFLGPSYSVLMFNPRDVERVLGSSQLTKS
FT OEYSFLGRWLNEGLVNSGRKWRHRRKIITPAFHRIEIPY
FT VEIIFDROSLRLVEELALRISGOERINLGEAIIHLCALDA
FT (IN SHORT ISOFORM).
FT E-> D (IN STRAINS CAM-2, CAM-3, CAM-8,
FT CAM-12, CAM-41, CAM-44, CAM-48 AND
FT BERKELEY).
FT S-> I (IN STRAINS CAM-8, CAM-44, CAM-48
FT AND BERKELEY).
FT Y-> I (IN REF. 1).
FT E-> K (IN REF. 1).
FT AIVANVLRHYEVDFVGDSSEPPVLAELILRTKEPIMKVVR
FT E-> PSMWMCSTRLTLMATSFGTTRADRTYSAYOGPL
FT SSRG (IN REF. 1).
SQ SEQUENCE 512 AA; 58691 MW; F7B089734231A6A3 CRC64;

Query Match 12.4%; Score 324.5; DB 1; Length 512;
Best Local Similarity 24.9%; Pred. No. 3.1e-15;
Matches 132; Conservative 83; Mismatches 209; Indels 107; Gaps 20;

QY 5 ALLASALCTFVLPILLFLAIAIKLMDLYCVSGRDRSCALLPPTMGPFPGETLQVLIQR 64
   ||||| || || | | | |
Db 7 AILASAL--FVGLLLYHLKFKRLIDL-----ISYMPGPVLPVYG-----\ 44

QY 65 RKFLQMKRRKYGFIYKTHLFGRT-----VRMGADNVRRILLGDRLYSVHMPA 114
   : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db 45 -----HGHFICGKPRHEMWKKIFEMETYSKDQVLKWLGPVLNLMGNPK 90

QY 115 SVRTILGS-----GCLSNLHDSHKQKKVIMRAFSREALCYVPV 155
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 91 DVEVVLGTLRFNDKAGEYKALEPWLKEGLLVSRGRKWKHK--RKIITPAFHRIIDQVEV 149

QY 156 ITEEVGS-----SLEQ-WLSCGERGLLVPEVKRLMFRIAMRILLGCEPQLADGDSQ- 208
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150 F--EKSRDLRLRMEDRLKHGESGFSLYDWINLCTMDTICETAMGVSINAQSNADSEYV 207

QY 209 QIVEAF-----EEMTRNLFSLPIDVPFSGLYRG-MKARNLIHARIEQNI---RAKICGLR 259
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 208 QAVKTIISMVLHKRMFNILYRFDLYMLTPLARAEEKALNVLHOFTKIIYQREEL--IR 265

QY 260 ASEAGQCKD-----ALQLLIEHSWGERG-LDMQALKOSSTELLFGCHETTAS 307
   : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db 266 EGSSQESSNDADVAGAKRKMAFLDILLQSTVD--ERPLSNLDIREVDTFMECHDTSS 323

QY 308 AATSLITYLGLYPHYLVQVREELSKGLICKSNQDNKIDMEILEQLKYIGCVIKETLRIN 367
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 324 ALMFFFYNIATHPQAQKCFEEIRS---VVGNDKSTPVSYELLNQLHYVDLCYKETLRMY 380

QY 368 PVPYGGFRVALKTFELNGYQIPKGNVYISICDTHVAEIFTNKEEFPNDRFS-APHPED 426
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 381 PSVPLGRKVLDECEINGKLIPAGTINIGISPLYLGRREELFSEPNSEFKPERFDVYTTAEK 440

QY 427 ASRFSFIPEGGGLRSCVSGKEFAKILLKIFTVELARHCDWQLL---NGPPTM 474
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 441 LNPYAVIPFASAGPRNCIGQKFMLEIKAIIVANVLRHYEVDVFGDSSEPPVL 491

RESULT 14
CP84_ARATH STANDARD: PRT; 520 AA.
AC 042600;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 84A1 (FERULATE-5-HYDROXYLASE) (EC 1.14.-.-) (F5H).
GN CYP84A1 OR FAH1 OR AT4G36220 OR F23E13.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=96293440; PubMed=8692910;
RA Meyer K., Cusumano J.C., Somerville C.R., Chapple C.C.S.;
RT "Ferulate-5-hydroxylase from Arabidopsis thaliana defines a new
RL family of cytochrome P450-dependent monooxygenases.";
RN Proc. Natl. Acad. Sci. U.S.A. 93:6869-6874(1996).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Anson W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermayer B., Mache R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boulry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schuren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Deftoor E.,
RA Weltzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzner E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
RA Pettett A., Rajandream M.-A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sehon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Mux P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones C., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777(1999).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG ERECTA;
RX MEDLINE=99097044; PubMed=9880351;
RA Ruegger M., Meyer K., Cusumano J.C., Chapple C.;
RT "The regulation of ferulate-5-hydroxylase expression in Arabidopsis in
RL the context of sinapate ester biosynthesis.";
CC -!- PATHWAY: GENERAL PHENYLPROPANOID PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; U38416; AAC49389.1; -
DR EMBL; AL022141; CAA18128.1; -
DR EMBL; AF068574; AAD11580.1; -
DR EMBL; AL161589; CAB80293.1; -
DR InterPro; IPR001128; -
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; p450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BINDING 458 458 HEME (BY SIMILARITY).
SQ SEQUENCE 520 AA; 58720 MW; E812779AF5BF01BC CRC64;

Query Match 12.3%; Score 322; DB 1; Length 520;
Best Local Similarity 25.0%; Pred. No. 4.7e-15;
Matches 129; Conservative 90; Mismatches 198; Indels 98; Gaps 22;

QY 15 VLPLLLFLAIAIKLMDLYCVSGDRSCALPLPPGTMGFPFEGTLLQWLVLRKFLQMKRRK 74
DB 22 VVSLFIFISFI-----TRRRRPPYPGPRGWPILGNMLMMDQLTHRGLANLAKK 70

QY 75 YGFIYKTHLFGPPTVRMGADNVRRILLGDDRLVSVHWPASVRYTILSGCLSNL-HDSS- 132
DB 71 YGGLCHLRMGFLHMYAVSSPEVARQVLQVDSVFS-NRPATL-----AISLYTD RAD 122

QY 133 -----HKQRKKVIMRAFSREALCYVPV-----ITEEVGSSLEQWLSCGERGL 176
DB 123 MAFAHYGPFWRQMKVKVCMKVFYSRKRAESWASVRDEVCKMVRYSVCNVGKPINVGEO--- 179

QY 177 VYPEVKRLMFRIAMRILLGCEPQLAGDGDSEQQLVEAFEEEMTR--NLFSLPIDVPFSGLY 234
DB 180 IFALTRNITYRAFG--SACE-----KGQDEFIRILQEFESKLFGAFNVADFIPIYFGWI 230

QY 235 -----RGMKARN-----LIHARIEQINIRAKICGLRASEAG-----QGCKDALQLIEHS 278
DB 231 DPQINKRLVKARNLDLGFIDDIIDEHMKKE--NQNAVDDGDVDTDMVDLLAFYSEEA 289

QY 279 WERGERLDMQ-----ALKOSTELLFGGHETTASATSLITYGLYPHYLVQKVEREL 330
DB 290 KLVSETADLQNSIKLTRDNKAIIMDVMEGGTETVASAIEWALTTELLRSPEDLKRYOQEL 349

QY 331 -KSKGLLCKSNQDNKLDMELEQLKYIGCVIKETLRLNPPVPGGFRVALKTFELNGYQIP 389
DB 350 AEVVG L-----DRRVESDIEKLTLYLCKTKEITLRMHPPIPLLHETAEDTSIDGFPI 403

QY 390 KGMNVIVSICDTHDVAEFTNKEEFPDRFSAPHED--ASRFSFIPFGGLRSCVGEF 447
DB 404 KKSVMINAFAGRDPTSWTDPDFRSPRFLPEGVPDFKGSNFEFIPFGSGRRSCPGMQ- 462

QY 448 AKILKIFTVELA----RHC-DWQLNGPPTMKT 477
DB 463 ---LGLYALDLAVAHILHCTWKL---PDGMKPS 490

RESULT 15

CP51_PIG STANDARD; PRT; 503 AA.

AC 046420;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450 51 (EC 1.14.14.1) (CYPL1) (P450L1) (STEROL 14-ALPHA DEMETHYLASE) (LANOSTEROL 14-ALPHA DEMETHYLASE) (LDM) (P450-14DM).
GN CYP51.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=Liver;
RA Kojima M., Morozumi T., Hamasima N., Okamoto T.;
RT "Cloning of a pig lanosterol 14-demethylase cDNA."
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: CATALYSES C14-DEMETHYLATION OF LANOSTEROL; IT TRANSFORMS LANOSTEROL INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH + OXIDIZED FLAVOPROTEIN + H(2)O.
CC -I- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC -I- SUBCELLULAR LOCATION: MICROSOMAL (POTENTIAL).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AB009988; BAA24134.1; -
DR InterPro; IPR001128; -
DR InterPro; IPR002403; -
DR Pfam; PF00067; p450; 2.
DR PRINTS; PR00465; EP450IV.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Transmembrane; Heme; Cholesterol biosynthesis.
FT TRANSMEM 24 44 POTENTIAL.
FT BINDING 449 449 HEME (BY SIMILARITY).
SQ SEQUENCE 503 AA; 56866 MW; 0302949CE461AFD6 CRC64;

Query Match 12.3%; Score 321; DB 1; Length 503;
Best Local Similarity 23.9%; Pred. No. 5.2e-15;
Matches 121; Conservative 87; Mismatches 244; Indels 54; Gaps 14;

QY 6 LLASAL--CTFVLPL-LFLAIAIKLMDLYCVSGDRSCALPLPPGTMG-----PPFG 55
DB 26 LLSLLACAFLLLVLLFRQAI-----GHLAPLPAGAKSPPIYFSPPIFLG 72

QY 56 ETLQWLVQRKFLQMKRRKYGFIYKTHLFGPPTVRMGADNVRRIL--LGDDRLVSVHWP 113
DB 73 HATAFGKSPIEFLENAYEKGVFSEFTMGKTFYLLGSDAALLFNSKNEDLNAEDVYS 132

QY 114 ASVYTIILSGCLSNLHDSHQRKKVIMRAFSREALCYVPVITEEVGSSLEQWLSCGER 173
DB 133 RLTPVEGKVAVDPNPVFLQKKMLKSLGINAHFRQHVSIIEKETKEYFQSWGESGER 192

QY 174 GLLVPEVKRLMFRIAMRILLGCEPQLAGDGDSEQQLVEAFEEEMTRNL--FS-----LP 225
DB 193 NL--FEALSELIITLASHCHLGE-----IRSQLEKVAQLYADLDGGFSHAALLP 242

QY 226 IDVPFSGLYRGMKARNLIHARIEQINIRAKICGLRASEAGQCKDALQLLIEHSWGERL 285
DB 243 GMLPLPSFRR---RDRAREIKNIIFYKAIQKRQSE--EKIDILQTLIDSTYKDGRL 296

QY 286 DMQALKOSTELLFGGHETTASATSLITYGLYPHYLVQKVERELSKGLLCKSNQDNKL 345
DB 297 TDDEVAGMLIGLLAGQHTSSTTSAMMGFTLARDKTLQEKCYLEOKT---VCGEDLP-PL 352

QY 346 DMEILEQLKYIGCVIKETLRLNPPVPGGFRVALKTFELNGYQIPKGMNVIVSICDTHDVA 405
DB 353 TYDQDKDLNLDRCIKETLRLRPIMTMRMMAKFPQTIVAGYTIIPGHQVCVSPYVNRKL 412

QY 406 EIFTNKEEFPDRFSAPHEDASRFSFIPEGGLRSCVGEKFAKILLKIFYELARHCDW 465
DB 413 DSWVERLDENPDRIYLDNPAASGEKFAVYVPFGAGRHRCIGENFAVYQIKTIWSTMLRIYEF 472

QY 466 QLLNGP-PTMKTSPYVYPVDNLPARF 490
DB 473 DLIDGYPTVNYTMTHTPENPVIRY 498

Wed Nov 7 09:29:25 2001

us-09-668-482-4.rsp

Page 13

Search completed: November 6, 2001, 13:39:40
Job time: 243 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:38:25 ; Search time 115.92 Seconds
(without alignments)
567.250 Million cell updates/sec

Title: US-09-668-482-4
Perfect score: 2611
Sequence: 1 MGLPALLASALCTFVLPPLL.....PTVYPVDNLPAFTTHERGET 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2434	93.2	497	11 Q9R1F4	Q9R1F4 mus musculu
2	2030.5	77.8	492	13 Q9PUB4	Q9PUB4 gallus gall
3	1739.5	66.6	492	13 Q93323	Q93323 xenopus lae
4	1041.5	39.9	512	4 Q9NR63	Q9NR63 homo sapien
5	1029	39.4	525	4 Q9NP41	Q9NP41 homo sapien
6	663.5	25.4	444	2 Q59990	Q59990 synechocyst
7	565	21.6	477	10 Q9LWY7	Q9LWY7 arabidopsis
8	535	20.5	107	13 Q9PUG2	Q9PUG2 gallus gall
9	519.5	19.9	485	10 Q9SJH2	Q9SJH2 arabidopsis
10	517.5	19.8	463	10 Q9FH76	Q9FH76 arabidopsis
11	512	19.6	457	10 Q65624	Q65624 arabidopsis
12	492.5	18.9	482	10 Q81077	Q81077 arabidopsis
13	450.5	17.3	465	10 Q9FMA5	Q9FMA5 arabidopsis
14	439	16.8	513	10 Q64989	Q64989 arabidopsis
15	439	16.8	513	10 Q9SCQ9	Q9SCQ9 arabidopsis
16	436	16.7	443	10 Q9LJK2	Q9LJK2 arabidopsis
17	434	16.6	489	10 Q9ZV72	Q9ZV72 arabidopsis
18	426	16.3	474	10 Q9LKH7	Q9LKH7 phaseolus a
19	426	16.3	478	10 Q9LN73	Q9LN73 arabidopsis

20	417.5	16.0	496	10 Q9FOY4	Q9fgy4 cucurbita m
21	411.5	15.8	465	10 Q9LH81	Q9lh81 arabidopsis
22	409	15.7	504	10 Q9SNG3	Q9sng3 oryza sativ
23	406	15.5	457	10 Q9M066	Q9m066 arabidopsis
24	406	15.5	524	10 Q23242	Q23242 arabidopsis
25	391.5	15.0	464	10 Q9LIC5	Q9lic5 arabidopsis
26	377	14.4	475	2 Q9X7G9	Q9x7g9 myxococcus
27	348.5	13.3	511	10 Q9SWR1	Q9swr1 liquidambar
28	347	13.3	486	11 Q9JUY3	Q9jiy3 mus musculu
29	347	13.3	503	11 Q9JIP8	Q9jip8 mus musculu
30	344	13.2	318	10 Q9LVY3	Q9lvY3 arabidopsis
31	338	12.9	521	10 Q9XFM2	Q9xfm2 lycopersico
32	335.5	12.8	513	10 Q9FVB8	Q9fvb8 brassica na
33	334	12.8	518	10 Q9FT38	Q9ft38 arabidopsis
34	333.5	12.8	512	10 Q9FX29	Q9fx29 arabidopsis
35	333	12.8	520	10 Q9FVC0	Q9fvc0 brassica na
36	332	12.7	520	10 Q9FVB9	Q9fvb9 brassica na
37	328.5	12.6	474	10 Q9SHY7	Q9shy7 arabidopsis
38	327	12.5	498	5 Q16806	Q16806 drosophila
39	326	12.5	464	10 Q04949	Q04949 arabidopsis
40	318.5	12.2	520	10 Q48786	Q48786 arabidopsis
41	318	12.2	503	5 Q21424	Q21424 caenorhabdi
42	317.5	12.2	735	10 Q9LGI7	Q9lgi7 arabidopsis
43	313.5	12.0	488	10 Q9SAA9	Q9saa9 arabidopsis
44	313.5	12.0	513	10 Q9SBP8	Q9sbp8 populus trl
45	312.5	12.0	491	5 Q9NGK3	Q9ngk3 tribolium c

ALIGNMENTS

RESULT 1
Q9R1F4 PRELIMINARY: PRT; 497 AA.
ID Q9R1F4;
AC Q9R1F4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CYTOCHROME P450 RETINOIC ACID METABOLIZING ENZYME P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TOOTH;
RX PubMed=11063033;
RA Paine C.T., Paine M.L., Snead M.L.;
RT "Identification of tuftelin- and amelogenin-interacting proteins using
the yeast two-hybrid system."
RL Connect. Tissue Res. 38:257-267(1998).
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
-!- OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AF115769; AADI7217.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 497 AA; 56191 MW; 67CB167A997842C1 CRC64;

Query Match 93.2%; Score 2434; DB 11; Length 497;
Best Local Similarity 93.2%; Pred. No. 1.4e-190;
Matches 463; Conservative 15; Mismatches 19; Indels 0; Gaps 0;
QY 1 MGLPALASALCTFVLPPLFLAIAIKLMDLYCVSGRDRSCALPLPGTMGFPFGETLQM 60
Db 1 MGLPALATATCTFVLPPLFLAALKLMDLYCVSSRRDRSCALPLPGTMGFPFGETLQM 60

Db 1 MDLYTLTSLA LCTLALPLLLLTAAKLWEVYCLRRKDAACANPLPGTMGLPFFGETLQM 60
QY 61 VLQRRKFLQMKRRKYGFIYKTHLEGRPTVRVMGADNVRRILLGDDRLVSVHWPASVRTIL 120
Db 61 VLQRRRFLQVRSQYGRITKTHLFGSPTVRVTGAENVROILMGEHKLVSVHWPASVRTIL 120
QY 121 GSGCLSNLHDSHKKQKKVIMRAFSREALECYVPVITEEVGSSLEQWLSCGERGLLVYPE 180
Db 121 GAGCLSNLHDNEHKYTKKVIQAQFSREALANYVPOMEVEYVCSVNLWLQSGP-CVLYYPA 179
QY 181 VKRLMFRIMRILLGCEPQLAGDGDSEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKAR 240
Db 180 IKRMFRIMARILLGCDPQRM-DREQEETLLEAFEEEMSRNLFSLPIDVPFSGLYRGLRAR 238
QY 241 NLIHARIEQNIRAKICGLRASAGGCKDALQLLIEHSWGERLDMOALKQOSTELLFG 300
Db 239 NLIHARIEQENIKEKL----QREDEHCKDALQLLIDYSRRNGEPINLQALKESATELLFG 294
QY 301 GHETTSASATSLITYLGLYPHYLQKVREELKSKGLL-CKSNQDNKLDMEILEQLYIGCV 359
Db 295 GHGTTASATSLTSFLALHKVDLEKVRKELETQGLSTKPEEKELSTIEVLQQLKATYSCV 354
QY 360 IKETLRNPVPVGGFRVALKTEFELNGYQIPKGNVITYSICDTHDVAEIFTNKEEFNPRF 419
Db 355 IKETLRISPPVAGGFVALKTFVINGYQIPKGNVITYSIADTHGEADLEPDTDKFNPRF 414
QY 420 SAPHPEDASRFSFIPEGGLRSCVCKEFAKILKIFTVELARHCDWQLNGPPTMKTSPT 479
Db 415 LTPLPDRSSRFEGFIPEGGVRCIGCKEFAKILKVFVVELCRNCDWELLNGSPAMTSPIT 474
QY 480 VYPPVDNLPARFTFHGEI 497
Db 475 ICPVDNLPAKFKPFSSSI 492

RESULT 4
Q9NR63 PRELIMINARY; PRT; 512 AA.
AC Q9NR63; 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME P450 RETINOID METABOLIZING PROTEIN P450RAI-2.
GN CYP26B1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20300913; PubMed=10823918;
RA White J.A., Ramshaw H., Taimi M., Stangle W., Zhang A., Everingham S.,
RA Creighton S., Tam S.-P., Jones G., Petkovich M.;
RT "Identification of the human cytochrome P450, P450RAI-2, which is
RT predominantly expressed in the adult cerebellum and is responsible for
RT all-trans-retinoic acid metabolism.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:6403-6408(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF252297; AAF76003.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; p450; 2.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 512 AA; 57512 MW; A06D1D9944E6726F CRC64;

Query Match 39.9%; Score 1041.5; DB 4; Length 512;

Best Local Similarity 44.2%; Pred. No. 8.4e-77;
Matches 219; Conservative 88; Mismatches 165; Indels 23; Gaps 9;
QY 7 LASALCTFVLPL----LFLAIAIKLMDLYCYSGDRSCALPLPPGTMGFPFFGETLQMYL 62
Db 8 LVSALATLAACLVSVTLILLAVSQQLMQLRMAATRDKSCKLPIPKSGMGFPLIGETGHWLL 67
QY 63 QRRKFLQMKRRKYGFIYKTHLEGRPTVRVMGADNVRRILLGDDRLVSVHWPASVRTILGS 122
Db 68 QSGSFQSSRREKYGWVEKTHLGRPLIRVTGAENVRKILMGEHHLVSTWMPRSTRMLLGP 127
QY 123 GCLSNLHDSHKKQKKVIMRAFSREALECYVPVITEEVGSSLEQWLSCGERGLLVYPEVK 182
Db 128 NTVSNSIGDIHRNKRKVFESKIFSHEALESYLPKIQLVITDLRAWSSHPE-AINYYQEAQ 186
QY 183 RLMFRIMRILLGCE-PQLAGDGDSEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRGMKARN 241
Db 187 KLTFRMAIRVLLGFSTPE-----EDLGHLEFYQQFVDNVFSLPVDLPFSGYRRGIQARQ 241
QY 242 LIHARIEQNIRAKICGLRASAGGCKDALQLLIEHSWGERLDMOALKQOSTELLFGG 301
Db 242 ILQGLEKAIREKL----QCTQCKDYLDALDLIESSKEHKEMTQELKDGTLELIFAA 297
QY 302 HETTSASATSLITYLGLYPHYLQKVREELKSKGLL---CKSNQDNKLDMEILEQLYIG 357
Db 298 YATTSASTSLIMQLKHPTVLEKLRDELRAHGILHSGCPC--EGTLRDLTSLGLRYLD 355
QY 358 CVIKETLRINPVPVGGFRVALKTEFELNGYQIPKGNVITYSICDTHDVAEIFTNKEEFNPD 417
Db 356 CVIKEMRLFTPISGGYRYLQTFELDGFQIPKGSVMYSIRDTHTAPAFKDVNVFDPD 415
QY 418 RFSAPHPEDA-SRFSFIPEGGLRSCVCKEFAKILKIFTVELARHCDWQLNGP-PTMK 475
Db 416 RFSQARSEDKDGFRHYLPRGGGVRTCLGKHLAKLKLAVELASTSRFELATRTPEPRIT 475
QY 476 TSPTVYPPVDNLPARF 490
Db 476 LVFVLPVDGLSVKF 490

RESULT 5
Q9NP41 PRELIMINARY; PRT; 525 AA.
AC Q9NP41; 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE WUGSC:H.NH0493L16.1 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1097-1108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Bourne S., Bauer C., Pape K., Jones T.;
RT "The sequence of Homo sapiens BAC clone Rp11-493L16.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
DR EMBL; AC007002; AAF65576.1; -.
Query Match 39.9%; Score 1041.5; DB 4; Length 512;

DR InterPro: IPR001128; -.
DR Pfam: PF00067; p450; 1.
DR PRINTS: PR00385; p450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 525 AA; 59124 MW; 71D47B6752A60315 CRC64;

Query Match 39.4%; Score 1029; DB 4; Length 525;
Best Local Similarity 43.1%; Pred. No. 9.1e-76;
Matches 219; Conservative 89; Mismatches 164; Indels 36; Gaps 10;

QY 7 LASALCTFVLP-----LLFLAAILKLDLYCVSGRDRSCALPLPGTGMFPFGETLQMWL 62
DB 8 LVSALATLACLVSVTLTLLAVSQQLQRLWATRDKSCKLPIPKSGMGFPLIGETGHWL 67
QY 63 QR-----RKFLQMKRRKYGFIYKTHLFGRPVRYMGADNVRRIILGDDRLVS 109
DB 68 QKCTLRTRVWLPQSGFGQSSRREKYGNVFKTHLGRPLRYTGAENVRKILMGEHLVS 127
QY 110 VHWPASVRTILGSGCLSNLHDSHKKQKKVIMRAFSSRALECYVPVITEVGSLSLEQWLS 169
DB 128 TEMPRSTRMLLPNTVNSISDILHRNKRKVFESKIFSHALESYLPKIQLVIODTLRAWSS 187
QY 170 CGERGLLVPEVKRLMFRIMRILLGCE-PQLAGDGSSEQOLVEAFEEEMTRNLFSLPIDV 228
DB 188 HPE-AINVYQEAQKLTFRMAIRVLLGFSIPE-----EDLGHLEFVYQGFVDNVFSLPYDL 241
QY 229 PFSGLYRGMKARNLIHARIEQNIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQ 288
DB 242 PFSGYRRGIQARQIILQKLEKAIREKL---QCTQKDYLDALDLIESSKEHGKEMTQ 297
QY 289 ALKQSTELLEFGHETTASATSLITYLGLYPHYLQKVRRELKSKGLL---CKSNQDNK 344
DB 298 ELKDGTLLELFAAYATITASTSLIMQLKHPVLEKLRDELRAHGILHSGGCPG--EGT 355
QY 345 LDMEILEQLKYIGCVIKETLRLNPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDV 404
DB 356 LRLDTLSGLRYLDVCIVKEWRLFTPISGYRTVLQTFELDGFQIPKGSVMYSIRDTHDT 415
QY 405 AEIFTNKEEFNDRFSAPHEDA-SRFSFIPEGGLRSCVGEKFAKILLKIFTVELARHC 463
DB 416 APVEKDVNVFDPDRFSQARSBDKDRFHYLPFGGVRCLGKHLAKLFLKVLAVELASTS 475
QY 464 DMQLLNGP-PTMKTSPTVYPVDNLPAF 490
DB 476 RFELATRTFPRITLVPVLAHPVDGLSVKF 503

RESULT 6
Q59990 PRELIMINARY; PRT; 444 AA.
ID Q59990; AC Q59990; DT 01-NOV-1996 (TREMBlrel. 01, Created) DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update) DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update) DE PUTATIVE CYTOCHROME P450 120. GN CYP120 OR CYP OR SLR0574. OS Synechocystis sp. (strain PCC 6803). OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis. OX NCBI_TaxId=1148; RN [1] RP SEQUENCE FROM N.A. RC STRAIN=PCC6803; RA Tabata S.; RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases. RN [2] RP SEQUENCE FROM N.A. RX MEDLINE=96127529; PubMed=8590279; RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N., RA Sugiyura M., Tabata S.; RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugiyura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.; RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: D64003; BAA10496.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; p450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Hypothetical protein; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 444 AA; 50578 MW; 8F62A9ED3B54BDC CRC64;

Query Match 25.4%; Score 663.5; DB 2; Length 444;
Best Local Similarity 33.7%; Pred. No. 5.2e-46;
Matches 152; Conservative 87; Mismatches 183; Indels 29; Gaps 8;

QY 41 ALPLPPTMGFPFGETLQMWLQRRKFLQMKRRKYGFIYKTHLFGRPVRYMGADNVRRI 100
DB 10 SLPIPRGDFGLPWLGETLNF-LNDGDFGKKRQQGFPIFKTRLEGKNVIFISGALANRL 68
QY 101 LLDGRLVSVHWPASVRTILGSGCLSNLHDSHKKQKKVIMRAFSSRALECYVPVITEEV 160
DB 69 FTKQETFPQATWPLSTRILLGPNALATQMGELHRSRKKILYQAFPLRTLDSTLPPKMDGIV 128
QY 161 GSSLEQWLSCGERGLLVPEVKRLMFRIMRILLGCEPQLAGDGSSEQOLVEAFEEEMTRN 220
DB 129 QGYLEQWKGANE--VIWYPLQRRMTFDVAATLFMGEKV-----SQNPOLFPPFETYIQG 180
QY 221 LFSLPIDVPFSGLYRGMKARNLIHARIEQNIRAKICGLRASEAGGCKDALQLLIEHSWE 280
DB 181 LFSLPIDLPNTLFGKSQARALLAELEKIIKAR-----QQQPPSEEDALGILLARD 234
QY 281 RGERLDMQALKQSTELLEFGHETTASATSLITYLGLYPHYLQKVRRELKSKGLCKSN 340
DB 235 NNQPLSLPELKQDILLFLFAGHETLTSSLCFLLGQHSDIRERVQEQN-----KLQ 288
QY 341 QDNKLDMEILEQLKYIGCVIKETLRLNPVPGGFRVALKTFELNGYQIPKGMNVIYSICD 400
DB 289 LSGELTAETLKMPYLDQVLEVLRLIPVGGGFRRELIDQCQFGFHPKGLVSYQISQ 348
QY 401 THDYAEIFTNKEEFNDRF-----SAPHEDASRFSFIPEGGLRSCVGEKFAKILLKIFT 456
DB 349 THADPDLYPDPPEKFDPERFTPDGSATIHNP--FAHYPRGGGLRECLGKFARLKMLFA 405
QY 457 VELARHCDWQLLNGPP-TMKTSPTVYPVDNL 486
DB 406 TRLIQFDWTLPLPGQNLLEYVTPSPRPKDNL 436

RESULT 7
Q9LVY7 PRELIMINARY; PRT; 477 AA.
ID Q9LVY7; AC Q9LVY7; DT 01-OCT-2000 (TREMBlrel. 15, Created) DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update) DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update) DE CYTOCHROME P450-LIKE. GN Arabidopsis thaliana (Mouse-ear cress). OS Arabidopsis; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; OC Eukaryota; eudicotyledons; core eudicots; Rosidae; eurosids II; OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20181125; Pubmed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:31-63(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AB018112; BAA96885.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450.1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 477 AA; 54850 MW; 304B4B2C4970E405 CRC64;

Query Match 21.6%; Score 565; DB 10; Length 477;
Best Local Similarity 31.2%; Pred. No. 6.4e-38;
Matches 154; Conservative 82; Mismatches 209; Indels 48; Gaps 15;

OY 18 LLLFLAIKLMDLYCVSGRDSCLP-LPPGTMGFPGFTLQMLQRR-----KFLQMK 71
Db 7 ILLFLSILLSLLLLRKHLHSFYPNLPBGNTGLPLIGSFSLSAGRQGHPEKFTDR 66
OY 72 RRRY----GFIYKTHLFGRPYRWGADNVRRIILGDRLYSVHWPASVRTILSGCLSN 127
Db 67 VRRFSSSSSCVEKTHLFGSPFAVVTGASGNKFLFTNENKLYVSWMPDSVKNKIFPSSMQTS 126
OY 128 LHDSSHQRKKVIMRAFSR-EALECYVPVITEEVGSSLE-QWLSGGERGLLVPEVKRLM 185
Db 127 SKEARKLR--MLLSQFMKPEALRRYGVMDLTAQRHFEFETW--ANQDVIVFPLTKFT 182
OY 186 FRIAMRILLGCEPQLAGDSEQLVEAFEEMTRNLFSLPIDVPFSGLYRGMKARNLIHA 245
Db 183 FSIACRSFLSME-----DPAVRYQLEEQFNTVAVGIFSIDLPGTRFNRAIKASRLRK 237
OY 246 RIEONIRAKICGLRASEAGGCKDALQLLIEHS-----WGERGLDMQALKOSTELLE 299
Db 238 EVSAIVRQKRELKAGKA-----LEHDLISHMLMIGETKD-EDLADKIIGLLI 286
OY 300 GGHETTASATSLITYLGLYPHVLQVREELKSKGLCKSNODNKLMEILEQLKYIGCV 359
Db 287 GGHDTASTYCTFVNVYLAEFPHVQRYLQE--QKEILKEKEKEGLRWEDIKMRYSWNV 344
OY 360 IKETLRINPPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNDRF 419
Db 345 ACEVMRIYPPPLSGTFREAIIDHFSFRGEYIIPKGMKLYWSATATFMNPDYFPEPEREFNRF 404
OY 420 --SAPHPEASRFSFIPFGGLRSCVCKEFAKILLKIFTVELARHCDWQ--LLNGPPTMKT 476
Db 405 EGSGPKP-----YTYVPFGGGPRMCPGCKEYARLEILIFMHNLVNRFKWEKVFPEENKIVV 459
OY 477 SPTVYPVDNLDPAR 489
Db 460 DPLPIPDKGLPIR 472

RESULT 8
O9PUG2 PRELIMINARY; PRT; 107 AA.
AC O9PUG2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYP26 (FRAGMENT).
GN CYP26.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Martinez-Ceballos E., Burdsal C.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL; AF185266; AAD56546.1; -.
DR HSSP; P14779; IBO7.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450.1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
FT NON_TER 1 107
SQ SEQUENCE 107 AA; 12009 MW; CA68BF8D598BC474 CRC64;

Query Match 20.5%; Score 535; DB 13; Length 107;
Best Local Similarity 89.7%; Pred. No. 2.1e-36;
Matches 96; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

OY 373 GFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNPDRESAPHEDASRFSF 432
Db 1 GFRVALKTLLENGYQIPKGMNVIYSICDTHDVAADLFTDKDEFNPDREMSPSPEDSSRFSF 60
OY 433 IPFGGLRSCVCKEFAKILLKIFTVELARHCDWQLNGPPTMKTSP 479
Db 61 IPFGGLRSCVCKEFAKILLKIFTVELARSCDWQLNGPPTMKTSP 107

RESULT 9
O9SJH2 PRELIMINARY; PRT; 485 AA.
AC O9SJH2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE CYTOCHROME P450.
GN AT2G42850.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; Pubmed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY

```
CC      SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL; AC006931; AAD21724.1; -.
DR DR InterPro; IPR001128; -.
DR Pfam; PF00067; p450; 2.
DR PRINTS; PR00385; p450.
KW PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electon transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KM Monooxygenase; Oxidoreductase.
SQ SEQUENCE   485 AA; 55405 MW;  08B0B16474620F82 CRC64;
```

Query Match

Best Local Similarity Matches 132; Conservative 86; Mismatches 201; Indels 47; Gaps	19.9%; Score 519.5; DB 10; Length 485; Pred. No. 3.4e-34;
--	--

```
OY    44 LPPGTMGFPFFGETLMVLQRRK-----FLQMKRKKYGFIYKTHLFGRPTVRVMGADV 97
       ||| | | :||| : ::          |: : ||| | | : | ||: | ||| :
Db     45 LLPGEMLPWIGETMTDFYKAQKSNRVEFEDEVNPRIKHGNIFKTTRIMGSPPIVNGAEAN 104
OY    98 RRLIGDDLVSVHWPAASYRTILSGCLSNLDSSHKQRKYIMRAF$REALECYVPVT 157
       . ||| : ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Db     105 RLISNEFSLVSSWPSSSYQLGMNCIMAKQGCKHRYLRGIIVANSLSYIGLESIPKLIC 164
OY    158 EEVGSSL-$WLSCGERGLIYPEVKRMFRIAMRIILGCPEOLAGDGDSQQLVFAEE 216
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     165 DTVKFHHETEMW-$GKEISLYRSAAKVLTFTVFECLEYIKYEIG-----MLEVFER 214
OY    217 MTRNLPSLPIDVPFGSLYRGMAKNLIHARIQNIRAKICGLURASEAOGCKDALQLITE 276
       : :||| : | | | | | | | | : | : | : | : | : | : | : | : | : |
Db     215 VLEGVALPVEP$SKAFRAKKARLEIETFVGVKVERE----REMEKEGAEKPNITLF- 269
OY    277 HSWMERELMQALKQSSTE-----LLFGHETTASATSLITYLGYPHYLOKVR 327
       || : : | | ||          |:| | ||| : | : | : | : | : | : |
Db     270 -----SRIVEELLKGVTIEEVVDNMVVLLVFAAHDTTSAMSMTFKMLAQHP---TCR 319
OY    328 ELKSKGLLKSN-$DNKLDMEILLEQLKIYIGCVIKEITRLNPVPVGGFRVALKTFELNG 385
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     320 DTLLQHAIQAANKAGEGYLTVEDYKKMYSMQVARETMRLSPPIFGSFRAVADIDYGG 379
OY    386 YQIPKGWNVIYSICTHDVAEIFTNKEEFNPD$APHEDEASRF$FIPEGGLRC$VCGK 445
       | ||||| : : : | | ||| : | : | | | | | | | | | | | | | | | | | |
Db     380 YTIPKGWKILTYYTGHTYNBEIFIQDPM$FDPTRFDKP---IQAYTYLPLPGGGPRLCAGH 435
OY    446 EFAKILLKITVELARHCMDQLNGRP$TMRKTSPTYVPYDNLPARFT 491
       : ||| : : | : | ||| : | : | : | : | : | : | : | : | : | : |
Db     436 QLAKISILVFMHFVNTGFDM$SVLYPDETISM$DP$LPPSIGMP$IKIS 481
```

```
RESULT 10
O9FH76 PRELIMINARY; PRT; 463 AA.
```

O9FH76

ID	AC	PRT:	463 AA.
Dt	01-MAR-2001 (TREMBLrel. 16, Created)		
Dt	01-MAR-2001 (TREMBLrel. 16, last sequence update)		
Dt	01-MAR-2001 (TREMBLrel. 16, last annotation update)		
De	CYTOCROME P450.		
Os	Arabidopsis thaliana (Mouse-eat cress).		
Oc	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
Oc	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;		
Ox	Brassicales; Brassicaceae; Arabidopsia.		
Rn	[1]		
Rp	SEQUENCE FROM N.A.		
Rc	STRAIN=COLUMBIA;		
Rx	MEDLINE=20181125; PubMed-10718197;		
Ra	Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,		
Ra	Tabata S.;		
Rt	"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence		
Rt	features of the regions of 3,076,755 bp covered by sixty pl and TAC		
Rl	clones.";		
Rl	DNA Res. 7:31-63(2000);		
Dr	EMBL; AB020744; BAB10255.1; -.		

SQ SEQUENCE 463 AA; 52366 MW; CCD17293F553F812 CRC64;
 Query Match 19.8%; Score 517.5; DB 10; Length 463;
 Best Local Similarity 30.0%; Pred. No. 4.7e-34;
 Matches 145; Conservative 82; Mismatches 220; Indels 37; Gaps 11;

```

QY      6 LLASALCTEVLPLLFLAIAIKIMDLVYSGRDRSCALPLDPGTMGFPFEGTLLQMLQ-R 64
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      8 LTLSAALF-LCLLRFIAGVR-----RSSSTKLPLPPGTMGYPYVGETFQLYSQDP 57
QY      65 RKFLQMKRRKYGFIYKTHLFGPRTVRMGADNVRRIILGDDRLVSVHWPASVRTILGSGC 124
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      58 NVEFAKQRRYGSVEKTHVLCPCVMISSPEAAKFVLYTKSHLFKPTFPASKERMLGQA 117
QY      125 LSNLHDSHKQKKVIMRAFSREALCEYVPVITEEVGSSLSQWLSCGERGLLVPEVKRL 184
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      118 IFFHOGDYHSLKRLVLRAFMPDAIRNMVPHITESIAQESLSNW--DGTQNTYQEMKTY 174
QY      185 MFIAMRILLGCEPQLAGDSDSEQQLVEAFEFEMTRNLFSLPIDVPFSGLYRGMKARNLIH 244
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      175 TENVALISILGKD-----EYVYREDLKRCCYILEKGYNSMPINLPGLTFHKAMKARKEL- 228
QY      245 ARIEQNIRAKICGLRASFAGQCGCKDALQLLIEHSWGERGLDMQALKOSSTELLFGCHET 304
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      229 AQLANL-----LSKRQNPSSHTDLGSMEDKAGLTDEQ-IADNIIGVIFARDT 279
QY      305 TASAATSLITYLGLYPHYLVQKVBELKSKGLCKSNQDNKIDMEILEQLKYGIVIKETL 364
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      280 TASVLTWLKYLADNPVYLEAVTEE-QMAIRKDKKGEESTLWEDTKMPLTYRVIOETL 337
QY      365 RLNPVPVGGERVALKTFELNGYQIPKGNVYISICDTHDVAEIFTNKEEFPNDRFS-APH 423
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      338 RAATILSFTFREAVEDEVYEGYLLIPKGWKVLPFRNIHNNADIFSDPGKFDPSREFEYAPK 397
QY      424 PEDASRFSFIPEGGGIRSCVCGKEFAKILKIFTELARHCDWQLLNGPPTMTKTSPTVYPV 483
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      398 PN-----TFMPFGSGIHSCPGNELAKLEISVLIHHLLTKYRWSIVGPSDGIQYGFALPQ 452
QY      484 DNL P 487
      : | |
Db      453 NGLP 456

RESULT 11
ID 065624 PRELIMINARY; PRT; 457 AA.
AC 065624;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYTOCHROME P450.
GN T18B16.200 OR AT4G19230.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-457 FROM N.A.
RA Van Der Schueren J., Chuang Y.J., Aert R., Defoor E., Robben J.,
RA Volckaert G., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
  
```

RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AL021687; CAA16713.1; -
 DR EMBL; AL161550; CAB78925.1; -
 DR HSSP; P33006; 1CPT.
 DR InterPro; IPR001128; -
 DR Pfam; PF00067; p450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 457 AA; 52436 MW; 6105FB7C181E4F07 CRC64;

Query Match	19.68;	Score 512;	DB 10;	Length 457;
Best Local Similarity	29.68;	Pred. No. 1.3e-33;		
Matches 144;	Conservative 84;	Mismatches 204;	Indels 54;	Gaps 13;

```

OY 18 LLLFLAIAIKLMDLYCVSGRDR--SCALPLRPGTMGFFFGFETLQWVLO-RRKFLQMKRRK 74
    | | : : | : | : : | | | | | | : | | : | : :
Db 8 LTLFAGSLFLYFLRCLISQRRFGSSKPLRPPGTMGNVYGETFQLYSQDPNVFFQSKOKR 67

OY 75 YGFIYKTHLFGRPYTRVMGADNVRRILLGDDRLVSVMHPASVRTILGSGCLSNLHDSHK 134
    | | : : | | : : : : : : : : | : | : : :
Db 68 YGSVEKTHVLGCPVCVMISSPEAKFVLVTKSHLFKPTFPASKERMLGKQALFFHQGDYHA 127

OY 135 QOKKVMIAFSSREALECYVPVYTEEVGSSLEQWLSGCEGCLLV--YPEVKRLMFRIAMRI 192
    : : : : | | : : | | | | | | : : : : | : | : : :
Db 128 KLRLVLRAFMPEsirRMVDPDIBSIAQDSLRSW-----EGTMINTYQEMKTYTFNVALLS 182

OY 193 LIGCEPOLAGDGDSEQOLVFAEFEMTRNLFSLPIDVPBSGLYRGMKARNLIHARIEQONIR 252
    : | : | : : : : : : : : : : : : : : : :
Db 183 IFGKDEVL----YREDLKRCYIILEKGYNSMPVNLPGTLEHKSMAKARKEL-----SQIL 232

OY 253 AKICGLRASAGGCGCKDALQLLIHSWGERLDMQALQOSSTELLFGGHETTASAATSL 312
    | : | | | | : : : : : : : : : : : : : : : :
Db 233 ARI---LSERRONGSSHNDLLGSEMGDKELTDEQ-IADNIGVIFAARDTASVMSWI 287

OY 313 ITYLGLYPHAVLOKVREELKSKGLCKSNQDNKLDMEILEOLKYIGCVIKETLRLNDPPVG 372
    : | | : | | : | | : : : : : : : : : : : : : :
Db 288 LKYLAEPNVYLEAVTEE--QMAIRKDKEEGESLTWGDTRKMBPLTSRVIQETLRVASILSF 345

OY 373 GFRVALKTFELNGYQIPKGNVNIYSICDTHDVAEIFTNKEEENPDRES-APHPEBASRES 431
    | | : : | | | | | | : : | : | : | : | : : :
Db 346 TEREAVEDVELEGYLLIPKGWVLPLEPRNIHISADIFSNPKRDPDSREFAVPKPN-----T 400

OY 432 FIPEGGGLRSCVGEKEFAKILLKI---FTVELARHCWQOLLNGBPMTKTSPTVYVPVDNLP 487
    | : | | | | | | | | : : | : | : | : | : : :
Db 401 FMPFGNGTHSCFQNELAKLEMSIMIHLLTKYRCVCVYLL-----IT 443

OY 488 ARFTHF 493
    | | | |
Db 444 FSETHF 449

```

RESULT	12	
081077		
ID	081077	PRELIMINARY; PRT; 482 AA.
AC	081077;	
DT	01-NOV-1998 (TREMBLrel. 08, Created)	
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)	
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)	
DE	PUTATIVE CYTOCHROME P450.	
GN	T9I4.17.	
OS	Arabidopsis thaliana (Mouse-ear cress).	
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;	
OC	Brassicales; Brassicaceae; Arabidopsis.	
OX	NCBI_TaxID=3702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	

RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
RA "Arabidopsis thaliana chromosome II BAC T914 genomic sequence.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AC005315; AAC33235.1; -.
DR InterPro; IPR001128; -.
DR pfam; PF00067; p450; 1.
DR PROSITE; P500086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 482 AA; 55175 MW; AB6A07AB2778DD3B CRC64;

Query Match	18.9%;	Score 492.5;	DB 10;	Length 482;
Best Local Similarity	28.4%;	Pred. No. 5.5e-32;		
Matches 141;	Conservative	90;	Mismatches 222;	Indels 43;
				Gaps 12;

[illegible]

RESULT	13
Q9FMA5	
ID	Q9FMA5
AC	Q9FMA5;
DT	01-MAR-2001 (TREMBLrel. 16, Created)
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE	CYTOCHROME P450.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC	Brassicales; Brassicaceae; Arabidopsis.
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=COLUMBIA;
RX	MEDLINE=98290546; PubMed=9628582;
RA	Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,

DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58867 MW; B1639BDD9A5D7C93 CRC64;

Query Match 16.8%; Score 439; DB 10; Length 513;
Best Local Similarity 27.0%; Pred. No. 1.4e-27;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 3 LPALLASALCTFVLPPLLLFLAIAIKLMDLYCVSGDRSCALPLPPGTMGFPFFGETL----- 58
Db 10 LPLLLPSTL-----LSLLFLILLK-----RRNRKTRFNLPPGKSGWPFLEGTIGYLK 57
QY 59 -QMYLQRRKFLQMKRRKRYGYTKTHLFGRPTVRVYGADNVRRIILGDDRLVSVHWPASVR 117
Db 58 PYTATTLGDFMQQHVSKYKGIYRSNLFGEPTIVSADAGLNRFILQNEGRLECSYPRSIG 117
QY 118 TILSGCLSNLHDSHQRKKVIMRAFSREALCEYVPVITEVGSS---LEQWLSGGER 173
Db 118 GILGKWSMLVLVGDMDHMRISINFLSHARLR---TILKDYERHTLFVLDWQ--NS 172
QY 174 GLVYPEVKRIMFRIAMRILGCEPQLAGDSE-QQLVEAFEEMTRNLFSLPIDVPESG 232
Db 173 IFSAQDEAKKFTFNMAKHIMSDP-----GEEETEQLKKEYVTFMKGVSAPLNLPGTA 227
QY 233 LYRGMKARNLIHARIEQNIRAKICGLRASEAGGCKDALQLLIEHSWE-----RGERL 285
Db 228 YHKALQSRATILKFIERKMERKLDIKEED-----QEEEEVKTEDEAEMSKSDHYRKQRT 282
QY 286 DMQ---ALKQS--STE-----LLFGHETTASATSLITYLGLYPHVLQKVRRELK 331
Db 283 DDDLIGWVLKHSNLSTEQILDLILSLFAGHETSSVAIALAIFFLQACPKAVEELREHHL 342
QY 332 SKGLCKSNQDNKLDMEILEQLKYGICVIKETRLINPPVPGGFRVALKTFELNGYQIIPKG 391
Db 343 EIARAKKELGESSELMWDYKRMDFTCVINETLRIGNVVRFLLHRKALKDVRYKGYDIPSG 402
QY 392 WNVVYSICDTHDVAEIFTNKEEFNPDREFSAPHPEASRFS-----FIPEGGLRSC 442
Db 403 WKVLPIYSAVHLDNSRYDQPNLFNFWRWQOQNNGASSSGSFSSTWGNNYMPFGGGPRLC 462
QY 443 VGKEFAKILLKIFTVELARHCDWQLNGRPTMTKSPTVYVVDNLP 487
Db 463 AGSELAKIEMAVFIHLLVLKFNWELAE-----DDKPFAPFVDFP 502

Search completed: November 6, 2001, 13:38:27
Job time: 255 sec

OM of: US-09-668-482-4 to: GenEmbl:* out_format : pfs
Date: Nov 6, 2001 3:55 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+pn.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09668482/runat_05112001_133353_10884/app_query.fasta_1.1680
-DB=GenEmbl -QFMT=fastap -SUFFIX=rg -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09668482@cgn1_1_8677 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-668-482-4
Query length: 497
Database: GenEmbl:*
Database sequences: 1344157
Database length: -85606004
Search time (sec): 5019.990000

sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pr4:AF005418	+ 2601.00	3750.00	1.2e-200	1743	AF005418 Homo sapiens retinoic
gb_r01:MMP450RA	+ 2437.00	3512.60	2.0e-187	1756	Y12657 M.musculus mRNA for P45
gb_r01:AF115769	+ 2434.00	3508.43	3.5e-187	1726	AF115769 Mus musculus cytochr
gb_htg21:AL358613	+ 2076.50	2946.30	7.1e-156	166337	AL358613 Homo sapiens chromo
gb_ov:AF199462	+ 2030.50	2926.02	9.6e-155	1479	AF199462 Gallus gallus retinoi
gb_ov:AF057566	+ 1739.50	2504.91	2.7e-131	1479	AF057566 Xenopus laevis retin
gb_ov:DRU68234	+ 1738.00	2502.74	3.6e-131	1479	U68234 Danio rerio all-trans-r
gb_pr5:AF252297	+ 1041.50	1484.02	2.0e-74	4445	AF252297 Homo sapiens cytochr
gb_ba3:SYCSLLE	+ 663.50	905.28	3.4e-42	113064	D64003 Synechocystis sp. PCG
gb_ov:AF185266	+ 535.00	776.78	5.0e-35	322	AF185266 Gallus gallus CYP26 (C
gb_p12:AF318211	+ 524.50	746.54	2.4e-33	1494	AF318211 Taxus cuspidata 5-alf
gb_p11:AC006931	+ 468.00	622.10	2.0e-26	116205	AC006931 Arabidopsis thalian
gb_p12:AF318500	+ 463.50	658.31	2.0e-28	1487	AF318500 Arabidopsis thaliana
gb_p14:SIU54770	+ 444.00	630.72	6.8e-27	1395	U54770 Solanum lycopersicum cy
gb_p12:AF318501	+ 436.50	617.70	3.6e-26	1740	AF318501 Arabidopsis thaliana
gb_pat1:AR074363	+ 433.50	614.13	5.7e-26	1608	AR074363 Sequence 1 from paten
gb_p12:ATCYP450R	+ 433.00	613.16	6.4e-26	1649	AF279252 Vigna radiata cytochr
gb_p12:AF279252	+ 417.50	590.33	1.2e-24	1718	AF212991 Cucurbita maxima ent-
gb_p12:AF212991	+ 406.00	572.52	1.2e-23	1934	AB008097 Arabidopsis thaliana
gb_ba3:MTCY22D7	+ 398.00	536.38	1.2e-21	31859	Z83866 Mycobacterium tubercu
gb_p11:AB018112	+ 398.50	526.43	4.4e-21	70475	AB018112 Arabidopsis thaliana
gb_p12:AF326277	+ 395.00	557.07	8.6e-23	1844	AF326277 Hordeum vulgare ent-k
gb_pr2:AC007002	+ 392.00	510.27	3.5e-20	140331	AC007002 Homo sapiens BAC cl
gb_p14:ZMU32579	+ 389.00	539.05	2.4e-22	1724	U32579 Zea mays DWARF3 (dwarf3
gb_ba3:MYTAH	+ 377.00	531.93	2.2e-21	1681	X87367 A.thaliana mRNA for cyt
gb_p13:ATT18B16	+ 377.00	492.22	3.5e-19	96574	AL021687 Arabidopsis thaliana
gb_p12:ATCHRIV50	+ 377.00	485.12	8.7e-19	199382	AL161550 Arabidopsis thalian
gb_p11:AB020744	+ 376.00	495.11	2.4e-19	62052	AB020744 Arabidopsis thaliana
gb_p12:AF139532	+ 348.50	487.58	6.3e-19	1830	AF166266 Mus musculus lanoster
gb_r01:AF166266	+ 347.00	487.69	2.1e-16	121125	AL022121 Mycobacterium tuber
gb_ba3:MTV025	+ 344.00	442.25	3.1e-18	2897	AB056457 Mus musculus Cyp4v3 m
gb_r01:AB056457	+ 341.50	475.22	3.1e-18	1960	AF150881 Lycopersicon esculent
gb_p12:AF150881	+ 338.00	473.99	3.6e-18	1835	AF214009 Brassica napus cytoch
gb_p12:AF214009	+ 335.50	471.02	5.3e-18	1880	AF214007 Brassica napus cytoch
gb_p12:AF214007	+ 333.00	467.16	8.7e-18	1884	AF214008 Brassica napus cytoch
gb_p12:AF214008	+ 332.00	465.69	1.1e-17	2356	M28655 Rabbit cytochrome P4501
gb_cm:RABCCYP4A5	+ 325.50	454.10	4.7e-17	2377	X57209 Rabbit cytochrome P4501
gb_cm:OCP450KD	+ 325.00	454.01	4.7e-17	2260	D55681 Rattus sp. mRNA for cyt
gb_r02:RATCP14DM	+ 325.00	453.78	4.9e-17	2383	U17697 Rattus norvegicus lanos
gb_r02:RNUI7697	+ 325.00	453.26	5.2e-17		

gb_in3:DMCTP450A	+ 324.50	457.20	3.1e-17	1481	Z23005 D.melanogaster cytoc
gb_pat1:AR085702	+ 322.00	451.46	6.5e-17	1838	AR085702 Sequence 1 from pa
gb_p13:ATU38416	+ 322.00	451.46	6.5e-17	1838	U38416 Arabidopsis thaliana
gb_cm:AB009988	+ 321.00	447.83	1.0e-16	2297	AB009988 Sus scrofa mRNA fo

seq_name: gb_pr4:AF005418

seq_documentation_block:

LOCUS AF005418 1743 bp mRNA PRI 15-DEC-1997
DEFINITION Homo sapiens retinoic acid hydroxylase mRNA, complete cds.
ACCESSION AF005418
VERSION AF005418.1 GI:2688845
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM human.
REFERENCE 1 (bases 1 to 1743)
AUTHORS White,J.A., Beckett-Jones,B., Guo,Y.D., Dilworth,F.J., Bonasoro,J., Jones,G. and Petkovich,M.
TITLE CDNA cloning of human retinoic acid-metabolizing enzyme (hp450RAI)
JOURNAL J. Biol. Chem. 272 (30), 18538-18541 (1997)
MEDLINE 97373542
REFERENCE 2 (bases 1 to 1743)
AUTHORS White,J.A., Beckett-Jones,B., Guo,Y., Dilworth,F.J., Bonasoro,J., Jones,G. and Petkovich,M.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1997) Cancer Research Labs, Queen's University, Botterell Hall, Rm 355, Kingston, Ont K7L 3N6, Canada

FEATURES
source 1..1743
/organism="Homo sapiens"
/db_xref="taxon:9606"
39..1532
/note="hp450RAI; cytochrome P450"
/codon_start=1
/product="retinoic acid hydroxylase"
/protein_id="AAB8881.1"
/db_xref="GI:2688846"
/translation="MGLPALLASALCTFVLPPLILFLAIIKLMDLYCVSGRDSALPL
PPTGMGFPEFGETLQWVLORRKFLQMKRRKYGFIKTHLFGPTVRVMGADNVRIIL
GDDRLVSVHMPASVRLTIGSGCLSLNHDSSHOKRKVIMRAFSALECYVPVITEEV
GSSLEQWLSGGERGLVPEYKRLMFLRIAMRILGCEPQLAGDSEDOIVAEFEEMT
RNLESLPIDVPFSGLYRGMKARNLIHARIEQNIKAKICGLRASEAGQCKDALQLLIE
HSWGERGRIDMALKQSTFELFGHETTSASATSLITYGLYPHVLQKREELSKG
LLCKSNQDNKIDMEILEOLKYLIGCVIKETLRNLNPPVPGGFRVALKTFELNGYQIPKGM
NVIIYSICDTHDAVEIFTNKEFNPDRLMPLHPEDASRESFIPFGGLRSCVGRFAKI
LKIFTEVLARHCDWQLNBPPTMTKTSPTVYPVDNLPAKFTFHGEI"

BASE COUNT	423 a	442 c	477 g	401 t
ORIGIN				
alignment_scores:				
Quality:	2601.00			497
Ratio:	5.255			0
Percent Similarity:	99.598			Percent Identity: 99.598

alignment_block:
US-09-668-482-4 x AF005418 ..
Align seg 1/1 to: AF005418 from: 1 to: 1743

1 MetcGlyleuProAlaIleuAlaSerAlaIleuCysThrPheValIleupr 17
|||||
39 ATGGGGCTCCCGGCGTGTGGCCAGTGGCTCTGCACCTTCGTGCTGCC 88
17 oleuleuIeuPheleuAlaIleuIleuIleuIleuIleuIleuIleuIleu 34
|||||
89 GCTGCTGCTCTTCTGCTGGCTGCATCAAGCTCTGGGACCTGTACTGCGTGA 138
34 erGlyArGAsParGSerCysAlaIleuProIeuProIleuGlyThrMetGly 50
|||||

139 GCGGCCGCGACCGCAGTTGTGCCCTCCCATTTGCCCCCGGAGCTATGGCC 188
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPhe 67
189 TTCCCTCTTCTTGGGAAACCTTGACAGATGGTACTGCACGCGAGGAAGTT 238
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
239 CCTGCAGATGAAGCGCAGAAATACGGCTTCATCTACAAGACGCATCTGT 288
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
289 TCGGGCGGCGCCACCGTACGGGTGATGGCGCGGACAATGTGGCGGCATC 338
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
339 TTGCTCGGAGACGACCGGCTGGTGTCCGTCACCTGGCCAGCGTCGGTCCG 388
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
389 CACCATTTCTGGGATCTGGCTGCCCTCTTAACCTGCACGACTCCTCGCACA 438
134 ySGlnArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
439 AGCAGCGCAAGAGGTGATTATGCGGGCTTCAGCCGCGAGGACACTCGAA 488
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluInTr 167
489 TGCTACGTCGCGGTGATCACCGAGGAAGTGGGCAGACGCTGAGCAGTG 538
167 PLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
539 GCTGAGCTGCGCGCAGCGCGGCTCCTGTCTACCCGAGGTGAAGCGCC 588
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
589 TCATGTTCCGAATCGCCATGCGCATCCTACTGGGCTGCGAACCCCACTG 638
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGluMe 217
639 GCGGGCGACGGGGACTCCGAGCAGCAGCTTGTGGAGGCTTCGAGGAAT 688
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
689 GACCCGCAATCTTCTCTCGCTGCCCATCGACGTGCCCTTCAGCGGGCTGT 738
234 yrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
739 ACCGGGCGCATGAAGGCGCGGAACCTCATTTACGCGCGCATCGAGCAGAAC 788
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
789 ATTGCGCCAAAGATCTGCGGGCTGGCGGCATCCGAGGCGGCGCAGGCTG 838
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
839 CAAAGACGCGCTGCAGCTGTTGATGACGACACTCGTGGAGAGGGGAGAGC 888
284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
889 GGCTGGACATGCAGGCACTAAAGCAATCTTCAACCGAACTCCTTTTGA 938
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGl 317
939 GGACACGAAACACGCGCAGTGCAGCCACATCTCTGATCACCTTACCTGGG 988
317 yLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerLysG 334
989 GCTTACCCACATGTTCTCCAGAAAGTGCAGAGAAGCTGAAGAGTAAGG 1038
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
1039 GTTACTTTGCAAGAGCAATCAACACAACAAGTTGGACATGGAATTTTG 1088

351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
1089 GAACAACCTTAATATACATCGGGGTGTGTATTAAAGAGACCCCTTCGACTGAA 1138
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
1139 TCCCCAGTTCACAGAGGGTTTCGGGTGTGCTGTGAAGACTTTTGAATTAA 1188
384 sngLYTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysasp 400
1189 ATGATACCAGATTCCCAAGGCGCTGCAATGTATTCTACAGTATCTGTGAT 1238
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
1239 ACTCATGATGTGGCAGAGATCTTCACCAACAAGAGAATTTAATCCTGA 1288
417 pArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
1289 CCGATTCAATGCTGCCCTCACCCAGAGAGATGCATCCAGTTCAAGTTCATTC 1338
434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1339 CATTTGGAGGAGGCGCTTAGGAGCTGTGTAGGCCAAGAATTTGCCAAAATT 1388
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
1389 CTTCCTCAAAATATTTACAGTGGAGCTGGCCAGCGATTGTGACTGGCAGCT 1438
467 uLeuAsnGlyProAlaProThrMetLysThrSerProThrValTyrProValA 484
1439 TCTAATGAGACTTCCTACATGAAACACAGTCCACCGTGTATCCGTGTGG 1488
484 sPAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
1489 ACAATCTCCCTGCAAGATTCAACCCATTTCCATTGGGGAATC 1529

seq_name: gb_rol:MMP450RA

seq documentation block:

LOCUS	MMP450RA	1756 bp	mRNA	ROD	08-JAN-1998
DEFINITION	M.musculus mRNA for P450RA protein.				
ACCESSION	Y12657				
VERSION	Y12657.1 GI:2765213				
KEYWORDS	P450RA gene.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 1756) Fujii,H., Sato,T., Kaneko,S., Gotoh,O., Fujii-Kuriyama,Y., Osawa,K., Kato,S. and Hamada,H.				
TITLE	Metabolic inactivation of retinoic acid by a novel P450 differentially expressed in developing mouse embryos				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1756) Hamada,H.				
AUTHORS	Direct Submission				
TITLE	Submitted (17-APR-1997) H. Hamada, Osaka University, Institute for Mol. & Cell. Biol., 1-3 Yamada-oka, Suita, Osaka, 565, JAPAN				
JOURNAL	location/Qualifiers				
FEATURES	1..1756				
source	/organism="Mus musculus"				
	/strain="C3H"				
	/db_xref="taxon:10090"				
	/cell_line="p19"				
gene	57..1550				
	/gene="P450RA"				
CDS	57..1550				
	/gene="P450RA"				
	/function="metabolises retinoic acid"				
	/codon_start=1				
	/product="P450RA protein"				

/protein_id="CAA73206.1"
/db_xref="GI:2765214"
/db_xref="SWISS-PROT:O55127"
/translation="MGLPALIASALCTFVLPLLFALAALKLMDLYCVSSRRSCALPL
PGTMGEFFEGETLQMWLQRRKFLQMKRRKYGFYKTHLFGPTRYVMGADNVRILL
GHEHLVSVHPASVRTILGACLSNLHDSHKQKKVIMQAFSREALOCYLVIAEEY
SSLEQWLSGGERGLVYPEVKRLMFRILMLGCEPGAEGEQLVEAEEMT
RNLFSLPIDVPFSGLYRGVARNLIHARIEENIRAKIRLQATEPDGGCKDALQLLIE
HSMERGERLDMOALKOSTELLFGHEHTASATSLITYGLYPHVLQKVEEIKSG
LICKSNQDNKLDMETLEQLKIGCVIKETRLNPNPVGGRVALKTFELNGYQIPKGM
NVYSICDTHDVADIFTNKEENPDREIVPHPEDASRFSFIPFGGLRSCVGRFAKI
LKIITVELARHCDWQLNGPPTMKTSPTVYPVDNLPAKFYFGDI"

BASE COUNT 408 a 434 c 494 g 420 t
ORIGIN

alignment_scores:
Quality: 2437.00 Length: 497
Ratio: 5.035 Gaps: 0
Percent Similarity: 97.384 Percent Identity: 93.360

alignment_block:
US-09-668-482-4 x MMP450RA ..

Align seg 1/1 to: MMP450RA from: 1 to: 1756

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
57 ATGGGGCTCCCGCGCTGCGCCAGTGCGCTCGACCTTCGTGCTGCC 106
17 oLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
|||||
107 GCTGCTGCTCTTCCTGCGCGCGCTCAAGCTCTGGGACCTGTACTGTGA 156
34 eRGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
||::|||
157 GCAGCGCGCATCGCAGCTGCGCCCTCCCTTGCCCCCGGTACCATGGGC 206
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
207 TTCCCATTTCTTTGGGAAACATTCAGATGGTGTTCAGCGGAGGAAGTT 256
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
257 TCTGCAGATGAAGCGCAGAAATACGGCTTCATCTACAAGACCATCTGT 306
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
307 TTGGGCGGCCACGGGTGCGGTGATGGCGCGGATATGTGGCGGCATC 356
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
357 TTGCTGGAGAGCACCGGTTGCTGCTGCACCTGCGCCCGCGCTGCGG 406
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
407 CACCATCCTGGGCGCTGGCTGCTCTCCCAACCTGCACGATTCCTCGACA 456
134 ySGlnArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
457 AGCAGCGAAAGAGGTGATTATGCAGCGCTTCAGCCGCGAGGCACATCCAG 506
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluInTr 167
|||||
507 TGCTACGTGCTCGTGATCGCTGAGGAAGTCACAGATTGTCTGGAGCAGTG 556
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
557 GCTAAGCTGGCGGCGAGCGCGCTCCTGTGTCTACCCCGAGGTGAAGCGCC 606
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
|||||
607 TCATGTTCGCGCATCGCATCGCATCTGTGGGCTGCGAGCGGGGTCCA 656

201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGluMe 217
|||||
657 GCGGCGCGCGGAGGAGCAGCAGCAGCTCGTGAGGCTTTCCAGAGAGAT 706
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
707 GACCCGCAATCTCTTCTCTCTCCATTCAGCTGCGCTTTAGCGGCTGT 756
234 yrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
757 ACCGGGCGTGAAGCGCGGAACCTTATACACGCGCGCATCGAGAGAAC 806
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
|||||
807 ATTCGCGCCAGATCCGCGGCTTCAGGCTACAGAGCCGGAATGGGGCTTG 856
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
|||||
857 CAAGACCGCGCTGCAGCTCCTGATTGAGCACTCGTGAGAGAGGGAGAGA 906
284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
907 GGCTGATATGCAGGCACTAAACAAATCGTCAACAGAGACTCCTTTGGT 956
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGl 317
|||||
957 GGTATGAACCTACAGCCAGTGCTGCAGATCACTGATCATTACTTACCTAGG 1006
317 yLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerLysG 334
|||||
1007 ACTCTACCCACATGTCCTCCAGAAAGTTGCGAAGAGATAAAGAGCAAG 1056
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
|||||
1057 GCTTACTTTGCAAGACCAATCAAGACACAAGTTAGACATGAAACTTTG 1106
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1107 GAACAGCTTAATATACATTGGGTGTGTCAATTAGAGAGACCCTGCGATTGAA 1156
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1157 TCCTCGGCTCCAGAGGGGTTTCGGGTGCTCTGAAGACTTTTGAGCTGA 1206
384 sngLyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
|||||
1207 ATGATATCCAGATCCCAAGGGCTGGAATGTTATTACAGTATCTGTGAC 1256
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1257 ACCCAGATGTGGCAGATATCTTCACTAACCAAGAGGAATTTAATCCCGA 1306
417 pArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||
1307 CCGCTTTATAGTGCCTCATCCAGAGGATGCTCCCGGTTCAAGCTTCATTC 1356
434 roPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1357 CATTTGAGAGAGGCCCTCGAGCTGTGTAGCAAAAGAGTTTGCAAAAATT 1406
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
|||||
1407 CTTCCTTAAGATATTACAGTGAGCTGGCTAGGCACATGTGATTGGCAGCT 1456
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1457 TCTAATTGACCTCTCAATGAGACAAAGCCCACTGTGTACCCCTGTGG 1506
484 spAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
|||||
1507 ACAATCTCCCTGCAAGATTCACTTCTCCAGGGAGAGATATC 1547

```
seq_name: gb_r01:AF115769

seq_documentation_block:
LOCUS      AF115769      1726 bp      mRNA      ROD      03-MAR-1999
DEFINITION Mus musculus cytochrome P450 retinoic acid metabolizing enzyme
            P450RA mRNA, complete cds.
ACCESSION  AF115769
VERSION    AF115769.1  GI:4324987
KEYWORDS
SOURCE     house mouse.
            Mus musculus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 1726)
            Paine,C.T., Paine,M.L. and Snead,M.L.
            Identification of amelogenin- and tuftelin-interacting proteins
            using the yeast two-hybrid system
            Connect. Tissue Res. 38, 257-267 (1998)
JOURNAL    2 (bases 1 to 1726)
            Paine,C.T., Paine,M.L. and Snead,M.L.
            Direct Submission
            Submitted (21-DEC-1998) CCMB, University of Southern California,
            2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
FEATURES
            source
            1..1726
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /tissue_type="tooth"
                /dev_stage="1 day old"
                27..1520
                /note="tuftelin-interacting protein 24 TIP-24"
                /codon_start=1
                /product="cytochrome P450 retinoic acid metabolizing
                enzyme P450RA"
                /protein_id="AAD17217.1"
                /db_xref="GI:4324988"
                /translation="MGLPALLATALCTFVLPLLLFLAALKLMDLYCVSSRDRSCALPL
                PPGTMGFPEFGETLQMYLQRRKFLQMKRRKYGFYKTHLFGPRTVRVMGADNVRILL
                GEHRLVSHWPASVRFILGAGCLSNLHDSHKQRKVINQAFSREALOCYVLVIAEEV
                SSCLQWLSCGERGLVYPEVKRLMFIAMRILLGCEPGDAGGEDQQLVEAFEEMT
                RNLESLPIDVPFSGLYRGVARNLIHARIEENIRAKIRLQATEIPDGGCKDALQLLIE
                HSWERGERLDMQALKOSSTELLFGHETTASAAATSLITLYLGLYPHLQKVREIISKG
                LLCKSNODNKLDMETLEQLYIGCVIKETLRNLNPVPGFGRVALKTFELNGYQIPRGW
                NVIYSICDTHVDADIFTNKEEFNPDRFIVPHPEDASRFSFIPFGGLRSVCVKEPAKI
                LKIFTELARHCDMQLNPPMTKTSPTVYVPVDNLPARFTYFQGD1"
BASE COUNT      405 a      428 c      476 g      417 t
ORIGIN
alignment_scores:
            Quality: 2434.00      Length: 497
            Ratio: 5.029      Gaps: 0
            Percent Similarity: 97.384      Percent Identity: 93.159
alignment_block:
US-09-668-482-4 x AF115769 ..
Align seg 1/1 to: AF115769 from: 1 to: 1726
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
27 ATGGGGCTCCCGCGGCTGCTGGCCACTGCGCTCTGCACCTTCGTGCTGCC 76
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
77 GCTGCTCTCTTCTTCTGGCGGCTCAAGCTCTGGGACCTGTACTGTGTGA 126
34 erGlyArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
127 GCAGCCGCATCGCAGCTGCGCCCTCTGCCCGCCCGGTACCATGGGC 176
51 pheProPheheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
```

```
177 TTCCCATTTCTTTGGGGAACAATTGCAGATGCTGCTTCACCGGAGGAAGTT 226
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
227 TCTGCAGATGAAGCGCAGGAATAACGGCTTCACTACAGAAGCATCTGT 276
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
277 TTGGGCGGCCACAGCTGCGGGTGATGGGCGCGGATATATGTGCGCGCATC 326
101 leuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
327 TTGCTGGAGAGACACCGGTTGGTGTCCGGTGCACCTGGCCCGCGTGGC 376
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
377 CACCATCCTGGCGGCTGGCTGCCCTCTCCAACCTGCACGATTCTCGCACA 426
134 ysgLInArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
427 AGCAGCGAAAGAAGTGATTATGCAGGCCCTTCAGCCGCGGAGGACCTCCAG 476
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
|||||
477 TGCTACGTGCTCGTGAATCGCTGAGGAAGTCAGCAGTTGTCTGGAGCAGTG 526
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
527 GCTAAGCTGCGGCGGCGGCGGCTCCTGCTCAACCCGAGGTGAAGCGCC 576
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
|||||
577 TCATGTTCCGCATGCCCATGCGCATCTGCTGGCGCTGCAGCCGGGTCCA 626
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGluMe 217
|||||
627 GCGGCGGCGGCGGAGAGACAGACAACAGCCTGTCGAGGCTTTCGAGAGAT 676
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
677 GACCCGCAATCTCTCTCTCTCCCATGTAGCGTCCCTTAGCGGCTGT 726
234 yrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
727 ACCGGGCGGTGAAGCGCGGGAACCTTATACAGCGCGGCATCGAGAGAAC 776
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
|||||
777 ATTCGGCGCCAAGATCCGCCGGCTTCAGGCTACAGAGCCGGATGGGGGTTG 826
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
|||||
827 CAAGACGCGCTGCAGCTCCTGATTGAGCACTCGTGAGAGAGGGAGAGA 876
284 rglLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
877 GGCTGATATGCAAGCACTAAACAATCGTCAACAGAGCTCCTTTGGT 926
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGl 317
|||||
927 GGTATGAATGAACACTACAGCCAGTGTGCGACATCACTGATCACTTACCTAG 976
317 yLeuTyrProHisValLeuGlnLysValArgGluLeuLysSerLysG 334
|||||
977 ACTCTACCCACATGTCTCTCCAGAAAGTTGAGAAAGATAAAGACGACAG 1026
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
|||||
1027 GCTTACTTTGCAAGACCAATCAAGACAAAGTTAGACATGGAACCTTTG 1076
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1077 GAACAGCTTAATAATACATTGGGTGTGTCAATTAAAGAGAGACCCTGCAATTGAA 1126
```

367 nProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1127 TCCTCCGGTCCAGAGGGTTTCGGGTGCTCTGAAGACTTTTGAGCTGA 1176
384 snGlyTyrGlnIleProLysGlyTyrPasnValIleTyrSerIleCysAsp 400
|||||
1177 ATGGATACCATGCCAGATCCCAAGGGCTGGAATGTTATTACAGTATCTGTGAC 1226
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1227 ACCCAGCATGTGGCAGATATCTTCACTAACAGAGGAATTATCCCGA 1276
417 PArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||
1277 CCGCTTATAGTGCCTCATCCAGAGGATGCTCCCGGTTCCAGCTTCATTC 1326
434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1327 CATTTGGAGGAGGCCCTTCGAGACTGTGTAGGCCAAGAGATTGCCAAAATT 1376
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnIle 467
|||||
1377 CTTCCTTAAGATATTACAGTGGAGCTGGCTAGGCACTGTGATTGGCAGCT 1426
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1427 TCTAATGGACCTCCTACAAAGAACAAGCCCCACCTGTGACCCCTGTGG 1476
484 spAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
|||||
1477 ACAATCTCCCTGCAGATTCACCTACTTCCAGGAGATATC 1517
seq_name: gb_htg21:AL358613

seq_documentation_block:
LOCUS AL358613 166337 bp DNA HTG 14-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-348J12, *** SEQUENCING IN
PROGRESS ***, 14 unordered pieces.
ACCESSION AL358613 GI:13660841
VERSION AL358613.8 GI:13660841
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166337)
AUTHORS Plumb, B.
TITLE Direct Submission
JOURNAL Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Apr 17, 2001 this sequence version replaced gi:13625010.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba348J12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 161818 bases at least Q40
Consensus quality: 163255 bases at least Q30
Consensus quality: 163957 bases at least Q20
Insert size: 165037; sum-of-contigs
Insert size: 176845; 5.5% error; agarose-fp
Quality coverage: 5.93x in Q20 bases; sum-of-contigs Quality
coverage: 5.81x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 32409: contig of 32409 bp in length
* 32410 32509: gap of 100 bp
* 32510 52913: contig of 20404 bp in length
* 52914 53013: gap of 100 bp
* 53014 55817: contig of 2804 bp in length
* 55818 55917: gap of 100 bp
* 55918 74512: contig of 18595 bp in length
* 74513 74612: gap of 100 bp
* 74613 83833: contig of 9221 bp in length
* 83834 83933: gap of 100 bp
* 83934 87001: contig of 3068 bp in length
* 87002 87101: gap of 100 bp
* 87102 97488: contig of 10387 bp in length
* 97489 97588: gap of 100 bp
* 97589 100583: contig of 2995 bp in length
* 100584 100683: gap of 100 bp
* 100684 105295: contig of 4612 bp in length
* 105296 105395: gap of 100 bp
* 105396 129702: contig of 24307 bp in length
* 129703 129802: gap of 100 bp
* 129803 132462: contig of 2660 bp in length
* 132463 132562: gap of 100 bp
* 132563 137099: contig of 4537 bp in length
* 137100 137199: gap of 100 bp
* 137200 158309: contig of 21110 bp in length
* 158310 158409: gap of 100 bp
* 158410 166337: contig of 7928 bp in length.
FEATURES
source
1..166337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-348J12"
/clone_lib="RPC1-11.2"
1..32409
/note="assembly_fragment:02593
fragment_chain:1
clone_end:T7
vector_side:left"
32510..52913
/note="assembly_fragment:01673
fragment_chain:1"
53014..55817
/note="assembly_fragment:01466
fragment_chain:2"
55918..74512
/note="assembly_fragment:01282
fragment_chain:2"
74613..83833
/note="assembly_fragment:02258
fragment_chain:2"
83934..87001
/note="assembly_fragment:00938
fragment_chain:2"
87102..97488
/note="assembly_fragment:01557
fragment_chain:2"
97589..100583
/note="assembly_fragment:00199
fragment_chain:2"
100684..105295
/note="assembly_fragment:01309"
105396..129702
/note="assembly_fragment:02028.0"
129803..132462
/note="assembly_fragment:02077"

```
misc_feature      132563..137099
                  /note="assembly_fragment:02005
                  fragment_chain:3"
misc_feature      137200..158309
                  /note="assembly_fragment:00282
                  fragment_chain:3"
misc_feature      158410..166337
                  /note="assembly_fragment:01837
                  fragment_chain:3
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      46344 a 33577 c 34990 g 50115 t 1311 others
ORIGIN
```

```
alignment_scores:
    Quality: 2076.50      Length: 1166
    Ratio: 4.221          Gaps: 13
    Percent Similarity: 42.196      Percent Identity: 41.938
```

alignment_block:
us-09-668-482-4 x AL358613 ..

Align seg 1/1 to: AL358613 from: 1 to: 166337

```
1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
54914 ATGGGGCTCCCGCGCTGCTGGCCAGTGCCTCTGCACCTTCGCTGCC 54963

17 OLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
54964 GCTGCTGCTTCTCTGCTGCTGCATCAAGCTTGCGGACCTGTACTGCGTGA 55013

34 erGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
55014 GCGGCCGCGACCGCAGTGTGCCCTCCCATTTGCCCGGAGCTATGGGC 55063

51 pheProPhePheGlyGluThrLeuGlnMetValLeu.Gln..... 63
|||||
55064 TTCCCCCTTTTGGGGAAACCTTGACAGATGTTACTGCAGGTAAAGGAGGG 55113

63 ..... 63

55114 TGGGGCGGAGAGGCTGCTTCCCGGAGCCGCGGCTGTGGCTTCT 55163

63 ..... 63

55164 GCTGAAGTCGGGGTAGCGCCCCCGGAGCATGCTATTGCGGTAGGAG 55213

63 ..... 63

55214 CAGGCTGGCGGAGACGCGGCGTCCCGGCGCCCTCATGCCACTTC 55263

64 .....ArgArgLysPheLeuGlnMetLysArg 72
|||||
55264 TCTCTCTCCGCTTCTCCACACAGCGGAGAGTTCCTGCAGATGAAGCGC 55313

73 ArgLysTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVa 89
|||||
55314 AGGAATACGGCTTCATCTACAGACGCATCTGTTGGGGCGGCCACCGT 55363

89 lArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAspa 106
|||||
55364 ACGGGTGAATGGCGCGGACAAATGTGCGGCGCATCTTGTCTCGAGAGCAC 55413

106 rGlyValSerValHisTrpProAlaSerValArgThrIleLeuGlySer 122
|||||
55414 GGCTGTGTGCTCCACTGGCCAGCGTGGTGGCACCAATTTGGGATCT 55463

123 GlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys.Lys. 138
|||||
55464 GGCTGCTCTTAACCTGCACGACTCTCGCACAAAGCAGCGCAAGAAAGT 55513
```

```
138 ..... 138

55514 GGGCGCAGAGCGCGGCTTGACAGGAGGGGACCCCATTTATGAGCG 55563

138 ..... 138

55564 GAATTCCGGCTGATGATGCTAGCGCGGGCTAGCAGCTTGAGGTGGCT 55613

138 ..... 138

55614 AGGACCTCTGCCAGCTCCAGGTTAGCTATACCACTCGAGAGTGCCA 55663

138 ..... 138

55664 TGTGTCGACAGACTGGGGGTGTCTGAAGGGGACGGCGTAGACGAGA 55713

139 .....ValIle 140

55714 GGGCGGATGAGGCTTTTAACGCTGTCCCTCTCGGAGCTCAGGTGAT 55763

140 eMetArgAlaPheSerArgGluAlaLeuGluCys..... 151
|||||
55764 TATCGGGGCTTCAGCGCGAGGACACTCGAATGCTACGTGCCGGTGATCA 55813

151 ..... 151

55814 CCGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 55863

151 ..... 151

55864 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 55913

152 ....TyrValProValIleThrGluValGlySerSerLeuGluIle 167
|||||
55914 NNNNTACGTGCCGGTGATCACCGAGGAAGTGGGCAGCAGCTGAGAGCAGT 55963

167 rPLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183
|||||
55964 GGCTGAGCTGCGGCGACCGCGCTCTGTTACCCGAGGTGAAGCGC 56013

184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnIle 200
|||||
56014 CTCATGTTCCGAATGCCATGCGCATCTTACTGGGCTGCGAACCCTCACT 56063

200 uAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGln 217
|||||
56064 GCGGGCGACGGGACTCCGAGCAGCAGCTTGTGAGGCCCTTCGAGGAAA 56113

217 eThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
|||||
56114 TGACCGCAATCTTCTCTGCTGCCATCGACGTGCCCTTCAGCGGCTG 56163

234 TyrArg..... 235
|||||
56164 TACCGGTAAGGGGGCAAAAGGGCTGCGGACTAGGGCGCGGACCTGG 56213

236 .....Gly.MetLysAlaArg 240
|||||
56214 GCGTGTGTCACCGCGCGGCTCTGCGCTCAGGGCATGAAGCGCGG 56263

241 AsnLeuIleHisAlaArgIleGluGlnAsnIleArgAlaLysIleCysGlu 257
|||||
56264 AACCTCATTCACGGCGGCATCGAGCAGACAAATTCGGCCCAAGATCTGCGG 56313

257 yLeuArgAlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeu 274
|||||
56314 GCTGGGGCATCCGAGCGGGCCAGGGGCTTCAAAAGACGCGCTGACGCTGT 56363

274 euIleGluHisSerTrpGluArgGlyGluArgLeuAspMetGln..... 288
|||||
56364 TGATCGAGCACTGTGTGGAGAGGGGAGAGCGGCTGGACATGACAGGTGAGT 56413

288 ..... 288
```


56414 AGCAGCTTCAGACGACGACTGCGAGTTGGTCCCTGGCTTCCAAGC 56463
288 288
56464 GCTGTTCCTGGGGCCCCAAAGCGCGCTGGGGCCCCAGCTTCTGCAG 56513
288 288
56514 TGGGCGCGGCTCAGACTACAGCTATGGAATCCCGAAGAGGCTGAGA 56563
288 288
56564 CACCCGGCTCAGAGAGCTGCGGAAGGGCTGCGGCGGAACCTGGAGCATC 56613
288 288
56614 CCCTAGCCTTCCAAGTTTCAAAGGAAAGTTGGAATTTGCAAAAATGTT 56663
288 288
56664 AATAAAGAACCTTGCAGATTTTATAAACTAAGACTTTAACTCAGAGATT 56713
288 288
56714 TCCGCGTAGAGCGGGGCTCGTACTCGCCTTACTGCTCCAGCTGAACATAAGG 56763
288 288
56764 GACGTTGCATTTTGTTTAAGATATTGCTTTCCTTGACTTTCGTGAGCA 56813
288 288
56814 AAACATTTAGCCCTTCTAGTCTTCCCTCCAGAACTCTCAGTTCGATTCTG 56863
288 288
56864 AGTAATCCCTTCTGTCAAAACGCGAGCAGACTTGTGAGAAATGTGGTCTCA 56913
289Ala..LeuLysGlnSerSerThrGluLeuPheGly 301
56914 CTGTATTCTTAGGCACTAAAGCAATCTTCAACCGAACTCCTCTTTGGAGG 56963
301 yHisGluThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuGly 318
|||||
56964 ACACGAAACCAAGCGGCGAGTGCAGCACATCTCTGATCATTACCTGGGGC 57013
318 eutyTProHisValLeuGlnLysValArgGluLeuLysSerLys... 333
|||||
57014 TCTACCCACATGTTCTCCAGAAAGTGCAGAGAAGCTGAAGAGTAAGGTA 57063
333 333
57064 GGGAGACTGGGTCTGGGGGTGCTCTTAATTAGCTTAGGAATTCAGCTGCT 57113
333 333
57114 CCTAGCCAACTTCCGAATAAGTACGTGTGCTGCCCTTCATGAGTATTTT 57163
333 333
57164 GAAAGTGCAAGGCCCAAGGCTGCGTGGGCCAGTGGGCAGAAATTAGCTTTG 57213
333 333
57214 TGAATAACAGAGATGAGTCTTGAGCTGTGAGCCCACTTGGCAGGGTTT 57263
333 333
57264 GACCTTGATTCCTACCCCTCCTCCACCTTTTGCTGAACCGTGAGATTTC 57313
333 333

57314 TCAGATAGGTTCCACTTCTTGAAATGGTGTGTCCAAAGGCATACATAGT 57363
333 333
57364 GTGAATAGCTGATTAGTGTGGGTGTGTGTGAGTGTGGGGTGGGGTT 57413
333 333
57414 TTTTGGGGGAACCTTGGTCTTCTAAGCTGTGAGCAGACATTCCTGGGA 57463
333 333
57464 TTGTAATAGATAGTGATTTGGCGAGGCAAAATGCCATTAGCTGCTGTT 57513
333 333
57514 TCCCCCTCCAGTCTCTCCCATCATGGGGCCCTTTGGGTTAGTCTCCAC 57563
333 333
57564 TTAAGCCCTGTTACGTCTGCTGGGCTGATTTTATTGAGCACAATAATA 57613
334Gly..LeuLeuCysLysSerAs 340
57614 CTGTTACACCTCTGTATGACTGTTTGTATAGGTTTACTTTGCAGAAGCAA 57663
340 nGlnAspAsnLysLeuAspMetGluIleLeuGluLysTyrIleG 357
|||||
57664 TCAAGACAACAAGTTGGACATGGAATTTTGGAAACAACACTTAATATCATCG 57713
357 lCysValIleLysGluThrLeuArgLeuAsnProProValProGlyGly 373
|||||
57714 GGTGTATTATTAGSAGAACCTTCGACTGAAATCCCCAGTTCAGAGGG 57763
374 pheArgValAlaLeuLysThrPheGluLeuAsn... 384
|||||
57764 TTTGGGTTGCTGTGAAGACTTTTGAATTAATGTAAATTAATTTCTCT 57813
384 384
57814 TCTTCTCCCTTTTGTGTGTGTTAAAACTCCTCTTCTTCCCTATGCT 57863
384 384
57864 GTGGCTGCAATCTTAAGCTTTTGATAATTGTCTGCCCTATATGAGAT 57913
384 384
57914 ATGTTTCAGCAACCTGATCCACCTCTCTTTCTACCTCTCCCTGCTT 57963
384 384
57964 TTAGCTCATATTTCGCCCAAGATATCAGTGAAGTCTTTTGTGTGA 58013
385Gly 385
58014 AAGTTAAATTCAGTTTGTTCAGCCCTTGGGGTTTCATTAATCTTTGCAGG 58063
386 ..TyrGlnIleProLysGlyTTrpAsnValIleTyrSerIleCysAspThr 401
|||||
58064 GATACCAAGATTCACCAAGGGCTGGAATGTTATCTACAGTATCTGTGATACT 58113
402 HisAspValAlaGluIlePheThrAsnLysGluLysPheAsnProAsp 418
|||||
58114 CATGATGTGCAGAGATCTTCACCAACAAGGAAGAAATTAACTCTGACCG 58163
418 gPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleProp 435
|||||
58164 ATTCAATGCTGCCTCACCCAGAGATGCATCCAGGTTCACTTCATTCAT 58213
435 heGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeu 451
|||||
58214 TTGAGGAGCCCTTAGGAGCTGTGTAGGCAAGAATTTGCAAAAATTTCTT 58263

452 LeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLeuLe 468
|||||
58264 CTCAAAATATTATACAGTGGAGCTGGCCAGCATGTGACTGGCAGCTTCT 58313
468 uAsnGlyProProThrmMetLysThrSerProThrValTyrProValAspA 485
|||||
58314 AAATGGACCTCCTACATGAAAAACAGTCCACCGTGATCTCTGTGACA 58363
485 snLeuProAlaArgPheThrHisPheHisGlyGluIle 497
|||||
58364 ATCTCCCTGCAAGATTACCCCATTTCCATGGGGGAAATC 58401

seq_name: gb_ov:AF199462

seq_documentation_block:

LOCUS AF199462 1479 bp mRNA VRT 21-NOV-1999
DEFINITION Gallus gallus retinoic acid degrading enzyme CYP26 mRNA, complete cds.

ACCESSION AF199462
VERSION AF199462.1 GI:6456740

KEYWORDS
SOURCE .

ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

REFERENCE 1 (bases 1 to 1479)
Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,
Jessel,T.M. and Eichele,G.

TITLE
JOURNAL Dev. Biol. (2000) In press
REFERENCE 2 (bases 1 to 1479)
Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,
Jessel,T.M. and Eichele,G.

TITLE
JOURNAL Submitted (28-OCT-1999) Max Planck Institute for Experimental
Endocrinology, Feodor-Lynen Strasse 7, Hannover 30625, Germany

FEATURES
source Location/Qualifiers
1..1479
/organism="Gallus gallus"
/db_xref="taxon:9031"
1..1479
/note="p450RAI"
/codon_start=1
/product="retinoic acid degrading enzyme CYP26"
/protein_id="AAF09250.1"
/db_xref="GI:6456741"

CDS
/translation="MGFSALVASALCTFLPLLEFLAIVRLMDLYCASGRDBSCPPLP
PGTMGLPFEGETLQAVLQRRKFLQMKRRKYGTYKTHLFGPTVRVMGAENVRHILL
GEHRLVSVQWPAWVRITLIGSGCLSNLHNGOHKHKVIMQAFSRDALOHVAVIQEEV
SACLAQWLGAQPCLLVPEVKRIMRIAMRILGFQROASPDEQQLVEAFEEIMIRN
LESLPIDVPFSGLYRGLRARNIIHAKIEENIRAKMARKEGKYKDALQILMEHTQGN
GEQLMQELKESATELLFGHETTASATSLIAFLGLHHDVLQKVRKELQKGLLSGP
NOEKOLNMEFLQIKVTGCVIKETLRLSPVPGEFRILKLTLELNGYQIPKGMNVIYS
ICDTHDVADLFTDKDEFNPDRFMSPPEDSSRFSPFGGGLRSCVGVKEFAKVLKIF
TVELARSCDMQLNGPPTMTKGPVIYVPDNLPAKFIGFSGOI"

BASE COUNT 302 a 438 c 427 g 312 t

ORIGIN

alignment_scores:

Quality: 2030.50 Length: 497
Ratio: 4.463 Gaps: 3
Percent Similarity: 91.549 Percent Identity: 78.873

alignment_block:
US-09-668-482-4 x AF199462 ..

Align seg 1/1 to: AF199462 from: 1 to: 1479

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17

|||||
1 ATGGGCTTCTCCGCTCTGCTGCGCCAGCGCCCTGTGCACCTTCCTGCTGCC 50
17 OLeuLeuLeuPheLeuAlaAlaIleLysLysLeuTrpAspLeuTyrCysValS 34
51 CCTGCTGCTCTTCTCTGCGCCGCGCTGCCGCTGTGGGACCTGTACTGCCCA 100
34 eGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
101 GCGGCGCGGACCCAGCTGTCCGCTGCCGTTGCCCGCGGACCATGGGG 150
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
151 CTGCCCTTCTTCGGGGAGACGCTGCAGATGTCCTGCAGCGCGGAAATT 200
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
201 CCTGCAGATGAAGCGCAGGAAATACGGCTTCATCTACAAGACTCACCCTCT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
251 TTGGGCGGCGCCACTGTGCGGGTGATGGGCGGTGAGAATGTGCGGCACATC 300
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
301 CTGCTGGGCGGACACCGCCCTGTGCTCCGTGCAGATGGCCCGCTCGGTGCG 350
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
351 CACCATCTTGGGCTCGGGCTGCTCTCCAACTTCCAAACGGGCGGACGACA 400
134 ySglnArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
401 AACACCGCAAAAGGTGATCATGCAGGCTTCTCCGGGATGCGCTGCAG 450
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGlnIleTr 167
451 CACTACGTGCTGTATCCCAAGAGGAGGTGATGCTGCTGCGTGCAGACAGTG 500
167 pLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
501 GCTGGGTGCGGGGCC...TGCCCTGCTGTGTATCCCGAGGTGAAGCGCC 547
184 eUmetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
548 TCATGTTCCGCATCGCCATGCGCATGCGCATGCTGCGCTTCCAGCCGCCAG 597
201 AlagLyaspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGluMe 217
598 GCCAGCCCC...GATGCTGAGCAGACGCTGCTGAGGCTTCGAGAGAT 644
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
645 GATCCGCAACCTCTTCTCCCTCCCATGCATGTGCTTTCAGTGGGCTCT 694
234 yTrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
695 ACCGGGGCTGAGGGCGCGCACATCATCCATGCCAAGATTGAGAGACAAC 744
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
745 ATCCGTGCCAAG.....ATGCCCGCGCAAGGAGCCGAGGGTGGCTA 785
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
786 CAAGGATGCACTGCAGCTGTTGATGAGCACACACAGGGTAATGAGAGAC 835
284 rGlyLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
836 AGCTCAACATGACAGAGAGCTGAAGAGAGTCTGCCACAGAGCTGCTTTTGA 885
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
|||||

886 GGTACGAAACCAACGCGCCAGTGTGCTGCTACATCACTGATTGCTTCCTAGG 935

317 yLeuTyrrProHisValleuGlInLysValArgGluLeuLysSerLysG 334
|||||:| |||||||||:|||||:| |||

936 GCTCCACCATGATGTTCTTGCAAGAAAGTGAGGAGAGCTGCAATTAAAGG 985

334 lYleuLeuCyLSysSerAsnGlnAspAsnLysIleuAspMetGluIleu 350
||||||| |||||:|:|:|:|:|:|:|:|

986 GGTACTGCTGTGGTCCCAACCAAGAAAAACAATTGAACATGGAGTCTTG 1035

351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||:| ||||| ||||| ||||| |||||:|

1036 GAGCAGCTGAAGTACACAGAGGCTGTGTATTCAAGGAACCCCTCAGCTGAG 1085

367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
:|||||:|:|:|:|:|:|:|:|:|:|:|

1086 CCCTCCTGTTCTCGAGGAGATTCCGAATTGCACCTCAGACCCCTTGAGCTGA 1135

384 snGlyTyrglnIleProLysGlyTrrPasnValIleTyrrSerIleCysasp 400
||||| ||||| ||||| ||||| ||||| |||||

1136 ATGGTTACCAGATCCCAAGGCTGGAATGTTATTACAGTATCTGGCAT 1185

401 ThrHisaspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
||||| |||||:|:|:|:|:|:|:|:|:|

1186 ACCCAGCATGTGGCCGATCTCTTACCAGACAAGAGCAGGTCAACCCGGA 1235

417 parGpHeSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
||||| :|:| |||||:|:| ||||| |||||

1236 CCGCTTCATGTCTCCGTTCCGCGAGAGACTCCTCTAGGTTCAGTTTCATTTC 1285

434 roPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
||||| ||||| ||||| ||||| ||||| |||||:|

1286 CTTTCGGTGGGGCTTGAGAGCTGCGTAGGCAAAAGATTGCCAAAAGTTC 1335

451 leuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnle 467
||||| ||||| ||||| ||||| ||||| |||||

1336 CTTCTAAAAATATTACGGTGGAGCTGGCGCGAGAGCTGTGACTGGCAGCT 1385

467 uIleuAsnGlyProProThrMetLysThrSerProThrValTyrrProValA 484
||||| ||||| ||||| ||||| ||||| |||||

1386 GCTAAATGGACCTCCTACCAATGAAGAAACGGGGCCCATAGTGTACCCCTGTGG 1435

484 spasnleuProAlaargPheThrHisPheHisGlyGluIle 497
||||| |||||:|:|:| ||||| |||||

1436 ACAATCTGCTGCCAAATTCATAGGTTTTCAGCGGCGCAAAATC 1476

seq_name: gbb_ov:AF057566

seq_documentation_block:					
LOCUS	AF057566	1479 bp	mRNA	VRT	10-MAR-1999
DEFINITION	Xenopus laevis retinoic acid converting enzyme mRNA, complete cds.				
ACCESSION	AF057566				
VERSION	AF057566.1	GI:3283171			

ORGANISM African clawed frog.
SOURCE xenopus laevis

REFERENCE AUTHORS TITLE	Abstract
1 (bases 1 to 1479) Hollemann, T., Chen, Y., Grunz, H. and Pieler, T. Regionalized metabolic activity establishes boundaries of retinoic	

JOURNAL	EMBO J. 17 (24), 7361-7372 (1998)
MEDLINE	99077803
REFERENCE	2 (bases 1 to 1479)
AUTHORS	Chen,J.L., Grunz,H., Panitz,F., Pieler,T. and Hollemann,T.
TITLE	Direct Submission
JOURNAL	Submitted (05-APR-1998) Developmental Biochemistry, University of

FEATURES	Location/Qualifiers
Source	1. .1479

```

CDS
    /organism="Xenopus laevis"
    /db_xref="taxon:8355"
    1..1479
    /note="RACE"
    /codon_start=1
    /product="retinoic acid converting enzyme"
    /protein_id="AAC25158.1"
    /db_xref="GI:3283172"
    /translation="MDLYTLTSLCTLALPLLLLTAAKLWEYVCURKDACANPL
    PPGTMGLPEFGETLQWVLRKRFQVRSQYGRITYKTHLFGSPVRYTGAENVROILM
    GEKLVSVHWPASVPTILGAGCLSNLHDNEHYTKVIAQAFSREALANYVPOMEERV
    RCSVNLWLQSGPCVLYIPAIKRMFRIAMRLLIGCDPQRMDEQEEYLLLEAFEEMSRN
    LPSLPIDVPFSGLYRGLRARNLIHAQIEENIKELQREPDHCKDALQLLIDYSRRNG
    EPINLQALKEASATELLFGHGHTASAATSLTSLALHKDYLEKVRKELETQGLSTKP
    EEKKELSTIEVLQQLKTYTSCVIKETLRISPVGAGFVALKTFVLNGYOIPKGMNVIYS
    IADTHGEADLPDPTDKFNPDRELTPLPDRSSRGFIPFGGAVRCCTGKEFAKILLKVF
    VVELCRNCDWELLNGSPAMTTSPIICPVDNLPKAFKPFSSSI"
BASE COUNT      350 a      422 c      384 g      323 t
ORIGIN

```

alignment_scores:	
Quality:	1739.50
Ratio:	4.055
Percent Similarity:	86.145
	Length: 498
	Gaps: 4
Percent Identity:	67.068

```
alignment_block:
US-09-668-482-4 x AF057566 .
```

Align seq 1/1 to: AF057566 from: 1 to: 1479

```

1 MetGlyLeuProAlaIleuLeuAlaSerAlaIleuCysThrPheValLeuPr 17
  ||| ||| ::|||:::|||||:::|||||
1 ATGGATCTGTACACTTGTCTCACCAGCGCTGTGTGACCCCTGGCAGTGGCC 50
17 OLeuLeuPheLeuAlaIleLeuLeuTrpAspLeuTyrCysValS 34
  |||||:::||||| |||||:::|||||:::
51 ATTGCTGCTCCTTCTCACCGCTGCCAAACTGTGGAGTCTACTGCTGA 100
34 eRgIyArGAsPaRgSerCysAlaLeuProLeuProProGlyThrMetGly 50
  ::||| ::||| ||||| ||||| |||||
101 GGCGCAAGATGCCGCGCTGAGCTAACCCACTGCCCCGGGCAACCATGGGA 150
51 pheProPhepheGlyGlyThrLeuGlnMetValLeuGlnArgArgLysPh 67
  ::|||:::|||||:::|||||:::|||||:::|||||
151 CTGCCCTTCTTCGAGAGACTCTGCAATGGTGCTACAGAGCGCAGGTT 200
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
  |||||:::|||||:::|||||:::|||||
201 CCTCCAAGTGAAGCGAAGCCAATACGGTGCATCTACAGACGCATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
  ||||| ||||| ||||| ||||| ::|||:::|||||
251 TCGGCAGCCCCACGGTGCAGCTACAGGCGCGAGAAAGTTGCACAGATC 300
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
  |||:::|||||:::|||||:::|||||
301 CTGATGGGGGAACACAAAGCTGGTTCGGTGCACCTGGCCGGCTGTGCG 350
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
  |||||:::|||||:::|||||:::|||||
351 CACCATCTTAGGGCGCGCTGTCTTCCAACCTGCACGACACAGCAGCACA 400
134 ySgInArGlySlyValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
  ||| ||||| ||||| ::|||:::|||||
401 AGTACACCAAGAAAGTGATGCGCGCAAGCTTCTCCCGGGAAGCCCTCGCT 450
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGlnGlnTr 167
  ||||| ||||| ::||| ||||| |||||
451 AATTACGTGCGCGCAGATGGAAGAAGAGTGAGATGCTCCGTGAACCTGTG 500
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
  |||||::: ||||| ::||| ||||| |||||

```


51 CGTTTACTCTTTCTCGCCGCGGTGAAGTTGTGGAGATGTTAATGATCC 100
34 ergLYArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
101 GACGAGTCGATCCGAACATGCAGAAAGTCTCTACCGCCAGGTACCATGGGC 150
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
151 TTGCCGTTCAATTGGAGAAACGCTCCAGCTGATCTCCAGAGAGGAAGTT 200
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
201 TCTGCGCATGAACGGCAGAAATACGGGTGCATCTACAAGACGCACCTCT 250
84 heGLyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
251 TCGGGAACCCGACTGTCTCAGGGTGATGGAGCTGATAATGTGAGGCAGATT 300
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
301 CTGCTGGCGGAACACAGACTGGTGTCTGTCTCAGTGGCCACGATCAGTGAG 350
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
351 AACCATCTGGGCTCTGACACCCCTCTCAATGTCATGAGTTCAACACACA 400
134 ysgLnaArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
401 AAAACAAGAAAAAGGCCATTATGAGGCGTTCTCTCGAAGATGCTTGAG 450
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
451 CACTACATTCCTCCGTATCCAGCAGAGAGTGAAAGCGCCATACAGGAATG 500
167 pleu.....SerCysGlyGluArgGlyLeuLeuValTyrProGluV 181
501 GCTGCAAAAAGACTCCTGC.....GTGCTGGTTTATCCAGAAA 538
181 alLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
539 TGAAGAAACTCATGTTCGGATAGCTATGAGAATCCTGCTTGTGAA 588
198 Pro...GlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAl 213
589 CCAGAGCAAAATAAAGACGGAC.....GAGCAAGAACTGGTGAAGC 629
213 apheGluGluMetThrArgAsnLeuPheSerLeuProIleAspValProp 230
630 TTTTGAGGAAATGATCAAAAATTTGTTCTCTCCCTGCAATCGACGTTCTT 679
230 heSerGlyLeuTyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArg 246
680 TCAGTGGTCTGTACAGGGGTTTGAGGGCACGCAATTTCATTCACTCCAAA 729
247 IleGluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAl 263
730 ATTGAGGAAACATCAGGAAGAAATTCAGAT...GACGACAATGAAGAAA 776
263 agLYGlnGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpG 280
777 CGAACAGAAATACAAAGACGCCCTTCAGCTGTGATCGAAGACAGAGAA 826
280 luArgGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGlu 296
827 GAAGTGACCAACCTTTTAGTTTGCAGCGCATGAAGAAGACGACTACAGAG 876
297 LeuLeuPheGlyGlyHisGluThrThrAlaSerAlaIleThrSerLeuI 313
877 CTCTATTGAGGATCATGAACCAACCGCCAGCAGCAGCTGCAACCTTCTGT 926
313 eThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluGluL 330
927 CATGTTTCTGGTCTGAACACAGAGAGTGGTGCAAGAGTCAAGAGGAGG 976

330 euLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAsp 346
977 TTCAGCAGAAAGTTGAATGGGCATGTATACACCTGGAAAGGCTTGAGT 1026
347 MetGluIleLeuGlnGlnLeuLysTyrIleGlyCysValIleLysGluTh 363
1027 ATGGAGCTGTTGGACCAGCTGAAGTACACTGGATGTGTGATTAAAGAGAC 1076
363 rLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysT 380
1077 TCTTGAATCAACCCCTCTGTTCCGAGAGATTCAAGATTCGACTCAAAA 1126
380 hrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyr 396
1127 CCTTGAATTGAATGTTACCAAAATTCCTAAAGATGGAACGTCAATTAC 1176
397 SerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysGluI 413
1177 AGCATCTGTGACACGACGATGTGGCCGACGCTTTCCAAACAAGAGGA 1226
413 upheAsnProAspArgPheSerAlaProHisProGluAspAlaSerArgp 430
1227 GTTCCAGCGGAGAGATTCAAGCAAAAGGCTGGAGGACGGGTCCAGGT 1276
430 heSerPheIleProPheGlyGlyLeuArgSerCysValGlyLysGlu 446
1277 TTAACCTACATCCCTTGGAGGAGATCCAGAGATGTGTGTGGCAAGAG 1326
447 pheAlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCy 463
1327 TTCGCCAAAGTGTACTCAAGATCTTTTAGTTGAGTTAACGCCAGCATTG 1376
463 sasPTpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrV 480
1377 CAATTGGATCTCTCAACAGCAGCCCGCAACAATGAAGAACAGGCCGACTA 1426
480 alTyrProValAspAsnLeuProAlaArgPheThrHisPhe 493
1427 TTTACCCAGTGAACAATCTCCCTACCAAAATTCAGTAGTTAT 1467

seq_name: gb_pr5:AF252297

seq_documentation_block: 4445 bp mRNA PRI 14-JUN-2000
LOCUS AF252297
DEFINITION Homo sapiens cytochrome P450 retinoid metabolizing protein
P450RAI-2 mRNA, complete cds.

ACCESSION AF252297
VERSION AF252297.1 GI:8515440

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS White,J.A., Ramshaw,H., Taimi,M., Stangle,W., Zhang,A.,
Everingham,S., Creighton,S., Tam,S.P., Jones,G. and Petkovich,M.
IDENTIFICATION of the human cytochrome P450, P450RAI-2, which is
predominantly expressed in the adult cerebellum and is responsible
for all-trans-retinoic acid metabolism
Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6403-6408 (2000)

TITLE
JOURNAL MEDLINE
2 (bases 1 to 4445)
REFERENCE
AUTHORS White,J.A., Ramshaw,H., Taimi,M., Stangle,W., Zhang,A.,
Everingham,S., Creighton,S., Tam,S.P., Jones,G. and Petkovich,M.
DIRECT SUBMISSION
JOURNAL Submitted (04-APR-2000) Cytochroma Inc., 116 Barrie Street,
Bioscience Complex, Suite 2424, Kingston, ONT K7L 3N6, Canada
location/Qualifiers

FEATURES
source
1. .4445
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .1539
CDS

1436 TCCTGCACCCGCTGGATGGCCTCAGCGTCAAGTTC 1470

seq_name: gb_ba3:SYCSLLE

seq_documentation_block: LOCUS SYCSLLE 113064 bp DNA BCT 13-FEB-1999 DEFINITION Synechocystis sp. PCC6803 complete genome, 22/27, 2755703-2868766. ACCESSION D64003 AB001339 VERSION D64003.1 GI:1001200 KEYWORDS

1-decarboxylase; cyanate lyase; cytochrome P450; dtdp-glucose 4,6-dehydratase; elongation factor EF-G; endo-1,4-beta-glucanase; ferredoxin-nitrite reductase; ferredoxin-thioredoxin reductase; catalytic chain; ferrochelatase; flavoprotein; low affinity sulfate transporter; methionine aminopeptidase; methyltransferase; molybdenum cofactor biosynthesis protein A; molybdenum cofactor biosynthesis protein C; molybdopterin biosynthesis MoeA; oligopeptide transport system permease protein; peptidyl-trna hydrolase; phosphoribosyl formylglycinamide cyclo-ligase; photosystem II P680 chlorophyll a apoprotein; polyA polymerase; endonuclease; replicative DNA helicase; succinate dehydrogenase iron-sulfur protein; trna-cys; trna-fmet(exon1); trna-fmet(exon2); Ycf34. Synechocystis sp. (strain:PCC6803) DNA.

Synechocystis sp. Bacteria; Cyanobacteria; Chroococcales; Synechocystis. 1 (bases 1 to 113064) Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N., Sugiyura,M. and Tabata,S. Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome DNA Res. 2 (4), 153-166 (1995) 96127529

2 (bases 1 to 113064) Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirosewa,M., Sugiyura,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K., Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A., Yamada,M., Yasuda,M. and Tabata,S. Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions DNA Res. 3 (3), 109-136 (1996) 97061201

3 (bases 1 to 113064) Tabata,S. Direct Submission Submitted (30-AUG-1995) to the DDBJ/EMBL/GenBank databases. Satoshi Tabata, Kazusa DNA Research Institute, Gene Structure 2: 1532-3 Yanauchino, Kisarazu, Chiba 292, Japan (E-mail:tabata@kazusa.or.jp, Tel:0438-52-3933, Fax:0438-52-3934) Potential protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis. location/Qualifiers 1. 113064

FEATURES source /organism="Synechocystis sp. " /strain="PCC6803" /db_xref="taxon:1143" complement(1. 3960) /note="ORF_ID:s110178" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAI0440.1" /db_xref="GI:1001201"

CDS

CDS

CDS

CDS

CDS

/translation="MSQQNYNWKREWCPRSGNISLRDGGYLDDEPNTKMGKACNPDIYS LEDIAIEPCLVLLGEPVGKTHELEKQRFTEETICEYSKYLGNLRTFTNLKEDLER DKTFSDWLEGHVHLHLFLDSLDEGLNVKSLAIGLVDEFKPKYKRLNHLRLTCR PLAFPSILIEEALKALMEESNFAIYELAPLRIDIVINAVEAEGSSDDFLKAIINORDIY PLAIKPISLGFLLNTYRRHNSQFPPEQKREYSLIEGCKCLCEVSKCRHALKETGNLD IDORLIVAARIAAVTIFANREFAWIDVEQGVPAEDVPROKLCFGEESNGREFEITQ EIVQEVLDPTGLESSRGLHRMGWAHQVYAEFAWYLTQRNIDLSKIKLLYSSADSTP KLIPOLHETAVWLASMRDNVFEQIEIKTDDVLLQTDIPDSELRATIVANLMOYEEE KLFDRGFCNVRNARAKLHSGLAEOQLKYICDSSKKEDTRRIAIDIAIECNLSLOGLD PROLTAELFQALTPSKRINLGMHRYFLESELVPQLDSEDLVIALDMLKTOGVRCFG HPEELGDAILLKAMEHVNVPGLVEFTOVALLOQKEHOKLISHSPSLQAQFOESLIG MNQRYALLENLVESIAEKDEDEDFYFLTSSMTEPIISSNDVDMULEKLQDSNHEKIQ RAMAKIIEEVFDRDRTKOISEIIKATONNSILSDVFYSYFTPIEDSEEVTOLRDHYL KMKNSWNPDLDPPEPKERVIDCLEQLESGLDSAWWRNLNREMTLLPOSKRYGHNKLD LTOLPGWKEAEQATKKRIIOGAIOYIOQOTNIDYSWIGTNTYNSSELYGKALQILIT EAPRTLDNFSTDIWKRWPTIILSVPSNFIDSNPRKNCRKLVSLSYHSPDEFLLTLK LIDKENODNNIEFVIDRLEMGCWDEKLENNLTEKVKDSOLKPNVFERKLEKLEQGAVOA KDIALSTIRNTNSEEDSEROKALFAIKALITYSDPDSWNVWEKITADVOGREAFE LAASCNLSGKFDLLEKQIADLYAWLWQYDPHYEDPDHRENVMAIYVSPRDNIAFFRN NLSQLRDKGTIEACAELELRIQELPHLPMLKETLTYAQQNMRRERWPEPKDKLLEF LIPEPSNLDLSNQSTIDRKTEKMEDEPKLINNISDSTINAPVGTSGVTSNNVTCSS PEKOKRFNWEILFVGIGIIVGVIATVGIFFNGILFNDVLKDSLNGSPSKI EQAMPKN" complement(4106. 4678) /note="ORF_ID:s110888" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAI0441.1" /db_xref="GI:1001202" /translation="MLAQPRKFSFDEYLAVEDDTDQLYELFNGELLAVPTESGLNFO IATFLLKFAEIVGYRRVGRGHLEFVHGEPKNEPDLTIRREHIAQLAKRNTIRLS MLPPLLVEVVSPELQREDDYIAKROQYQACGIPEYWIVDPOQTQVLYVELLAGCYO EYGNFEKVMAIQSPFPQLSLTVNEIFKDL" 4733. 4930 /note="ORF_ID:ssr1513" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAI0443.1" /db_xref="GI:1001203" /translation="MAIAKFEFPLQKSGRRLYTGFAPKESDITDVFPLOALTVEEI AAEYARELFAGDNYTDCLYPK" complement(4987. 6327) /note="ORF_ID:s110887" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAI0444.1" /db_xref="GI:1001204" /translation="MLSLRKRYALLMPSLWGLVIVGLLFFPPSPVWATESPIPLTDE QLEVGDSLDKFAFAATEAGDFVFAEKYWTLEIEKFPONPAVWSNRGNSRVSONKLDIEA IADFNQAIETLAPQTDPLYLNRGTLEAKGEFKAAIADYRVYLVNPEDAFAYNNRGNA EGGLEGNWTSALIEDFOQATAIAPNFAQANTALALAYELGQKTEAIIOTMRRLVKKYPMF ADMRAALTAVLWVVDGQGEAESNMWVAAGVDSRYQDLDMVKQIRRWPPSVISALDKFL SLS" 7468. 8031 /note="ORF_ID:s1r0895"

```

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAI10445.1"
/db_xref="GI:1001205"
/translation="MMSGKRLRSKSIDPSQLTAAQVIVSQVDALITDAVASEAGV
SKGGLLEFPTKEALVAGVQQLADRYETLHQELANDPAPDTPGHVYAYIRASALD
DOENELHFNLLAANTNPFLAPMRHFWQECHQIMASGLDPAVAATVIRLAMDGLWL
THLGFADLEDPLRSQVATATALKLAQL"
8165. 9550
/note="ORF_ID:slr0896"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAI10446.1"
/db_xref="GI:1001206"
/translation="MTDISMRFGMQAEREFLKLVAPLAGAOMAQAATGFVDVMGMGW
LGDVLAAGGLATMIFMAFMMTGVGLISGVSPLVAEYAGQTRIGQLTROGLWIVL
LISGMLFITHLNGVMHMSGQADPTILSDSYLVNMAWGLFPAIATFLRGCIIALS
QAPMLPLVIAATLTLENILGNVGLGFGKMGFPALGITGLVAIASISAHWMFLSLVYML
WKRPHOYSLIEDSLHLKPRILQQLNMGVPIGIAVLEYGLTTFSEMGALGTPIL
AAHQVVSQTVLVLEWVPLAMSYAATVYVGOWFGQOHWPQIROALVSTGLAVLEMLTA
GIALAYPQQIIGLYLDLNDPANGELNVGISIMKIAFGVLVDGLQRTANGVLOGLQ
DTRIPNVLGTIAYWGIGLTAASYLLGFYTPLSGAGVWIGTYIGLAASIAYLWRFQKVM
VRKRSLLPVAS"
9852. 13064

```

```

/note="ORF_ID:slr0897"
/codon_start=1
/transl_table=11
/product="endo-1,4-beta-glucanase"
/protein_id="BAA10447.1"
/db_xref="GI:1001207"
/translation="MNSILSPITVAVGGSLMWNGFTADFTITNTSTTSLSNWTYTFDT
VHKISGNPWGATFTSVDLNGCITRYLLTGAGWASSIPGGSVTIGFNGTQGTAGISNG
SLTSALLFTDVNANTGGTGGEHQHDPIPTDPVVDPTPNPGTGRFAANPAADIVG
ENPVAEDLDFGDVSVHNLILGKTSKVAIINPWMTPEFQIRGSFDDLTAANGFI
VQNEHLRODIDGVLWELSLGKARNSTVYVRSHHEYQOERIENENPDTMKLSFLYGT
RERLTVSDTPEGILLATOPTGQTLTVGTRSQLIPANLEFHHDVREDLEDAFGVT
VEQYTVVSRKTGLATPQGPAGQITDGHQSRPGNGLTPIGGEEFGGAGHSGHDGDI
DHGGMDHSGSDPGDGGGTPTNPDEPTPPTDPVSPSPFTVYVGGNWMNGETA
ELLTNVSGSTKLNSWSTFEDTYHKISGSPWATVQSTDLGGGITRYVVTGSEMAASIA
PGSSVYKFGNGTQGTALANGDALADLLELFGGIVGQSTGSGGNTGTPTPNPPTD
PIENPNPVDVPDPPAIANGDNHYEALQKSLFLEYEANRSGDLDDESTKRIWRGDSGL
RDGSDGIYEGGSGSNGLOAGLSLDLTGGYHDAGDHGKFGLPASTYLLTLLWGVEFTD
GYQTSGQMDLGLATIRWGTIDYLLKHAVTNANGETVEFVAQVGNASADHSLWSAPESQ
IARPMAWDTANKPGSDVAAGSAAALAAASILFREAGDFAYADELLYNAQSLRYAFDY
RGKYSDSIPEVRNYTNSWSGSEDELAYGAAWLSRAVNSAGDGSAVLQKALNITNSDI
GGLSRGWTTHNMWDASYGVAVILAQDTGNQAIKQDVTRLMDAVNNGTDGVQITDGGLRF
ISOWGSLRYAANTAMLAAYSDGLTDPGTKYAQLAQDTVDYILGSPRNASVMYVGYGN
DFEQQLPHRYAASGVGWDFGRNLEPNEHILLEGALVGGPTAANDFSINDSRDDYISNEVA
IDYNAGLTGALAWSVEQFGGNPLTAQUNLPGISAPQM"
13423..14931
/gene="nIra"
13423..14931
/gene="nIra"
/note="ORF_ID:slr0898"
/codon_start=1
/transl_table=11
/product="ferreredoxin-nitrite reductase"
/protein_id="BAA10448.1"

```

alignment_scores:		
Quality:	663.50	Length: 451
Ratio:	2.212	Gaps: 8
Percent Similarity:	66.519	Percent Identity: 33.703

```
alignment_block:
US-09-668-482-4 x SYCSLE ..
```

Align seg 1/1 to: SYCSLE from: 1 to: 113064

41 AlaLeuProLeuProProGlyThrMetGlyPhePheProPhePheGlyIuTh 57
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

[illegible]


```

/product-"5-alpha-taxadienol-10-beta-hydroxylase"
/protein_id="AAK00946.1"
/db_xref="GI:12656592"
/translation="MDSFIFLRISIGTKEGLESPAILSLTIAPLATIILLERYNH
RSSVKLPKGKIGFPLIGETIQLRLTLRSETPQKFFDDRLLKFGPVYMTSLIGHPTVVL
CGPAGNKLVLSNEDKLIVEMEGPKSFMKLIGEDSIVAKRGEDHRLRLTALFALGAOAL
ONYIGRMSSEIGHHENEKWKQKDEKVKVPLDVRGLLSIASTLFEDVANGHQOKLHL
LETLVGSLSVPLDPFGTRYRKGLQARIKLELIDELISLIRRRDLRSCIASDDODLLS
VLTFRDEKGNLSITDOGLIDNFSAMFHASDYTTVAPMALIFKLAYSNEYHEKVFQEQ
LEILGNKKEGEEISWKDLKSMKTYQWAOESLRMYPVFGIFRKATITDIHYDGYTIIPK
GWRVLCSPYTHLREYEFPEPEEFRRPSFEDEGRHVTPTYTVFGGGGRLTCPGWEFISK
LEILLFVHFEVKNSSYIPVDPNENEKVLSDPLPPLPANGFSIKLFPRS"

```

alignment_scores:		
Quality:	524.50	Length: 496
Ratio:	1.720	Gaps: 11
Percent Similarity:	61.492	Percent Identity: 28.427

```
alignment_block:
US-09-668-482-4 x AF318211
.
```

Align seg 1/1 to: AF318211 from: 1 to: 1494

[illegible]

198	ProGlnLeuAlaGlyAspGlyAspSerGlnGlnGlnLeuValGluAla	214
616GTAATGATGGACACCACAAGCAACTTCATCATCTTCT	656
214	eGlnGluMetThrArgAsnLeuPheSerLeuProIleAspValProPhe	231
657	GGAACTATTCTTGTGGAGATTGTTCAGTCCCGCTGACTTCACGAA	706
231	exGlyLeuTyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIle	247
707	CTCGTTATCGTAAAGGCTTCAGCGCGCTGAAGCTGATGAATCTTC	756
248	GluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaG	264
757	TCTCTCTAATAAACGACAGAGAGAGATCTGCGTTCAGGATAGCTTC	806
264	ylGlnGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGlu	281
807	TGATGAT..CAAGATCTACTGTGCGTGTGCTCACCTTCAGAGATGAAA	853
281	rgGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeu	297
854	AAGGGAACCTACGTGACAGACCAGGGGATCTTGACAACTTTTCTGTATG	903
298	LeuPheGlyGlyHisGlyThrThrAlaSerAlaAlaThrSerLeuIleTh	314
904	TTTCATGCTTCATATGACACCACCTGTTGCACCAATGGCTTGATATTAA	953
314	rTyrLeuGlyLeuTyrProHisValLeuGlnLysVal..ArgGlnGlu	330
954	GCTTCTATCTCCAAATCCTGAATACCATGAAAAAGTATTTCAGAGCAGT	1003
330	eulysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuasp	346
1004	TGGAA.....ATAATTGGCAATAAAAAGGAGGAAAGAAATCAGT	1044
347	MetGluIleLeuGlnLeuLysTyrIleGlyCysValIleLysGluTh	363
1045	TGGAAGGATTGAAATCTATGAATATATACATGGCAAGCAGTTCAGAATC	1094
363	rLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysT	380
1095	ACTACGAATGTACCCACCACAGTTTGTGAATATTTGTAAGCTATCACGTG	1144
380	hrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyr	396
1145	ATATTCAATTATGATGGGTATACAATTCAAAAGATGAGGGTTTATGT	1194
397	SerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysGluG	413
1195	TCACCTTATACTACACATCTGAGAGAGAGTACTTCCCTGAGCCTGAAGA	1244
413	upheAsnProAspArgPheSerAlaProHisProGluAspAlaSerArg.	429
1245	ATTCAAGCCCTTCAAGATTT.....GAGGATGAAGGCAGGC	1279
430PheSerPheIleProPheGlyGlyLeuArgSerCys	442
1280	ATGTGACTCCCTTACACATATGTACCATTGGAGGAGGCGCTGCCACATGT	1329
443	ValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrValGlu	459
1330	CCAGGATGGGAATTTCAAGATTTGAATAATTACTGTTTGTCCATCATTTT	1379
459	uAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrMetLysT	476
1380	CGTTAAAAATTTACAGCAGCTACATTCAGTTGATCCCAATGAAAAAGTTT	1429
476	hrSerProThrValTyrProValAspAsnLeuProAla	488
1430	TATCAGAT.....CCACTACTCTCTCCCTGCC	1458

seq_name: gb_pl1:AC006931

seq_documentation_block:
LOCUS AC006931 116205 bp DNA PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 231 of 255 of the
complete sequence. Sequence from clones F14N22, F7D19.
AC006931 AE002093
AC006931.5 GI:6598641
HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 116205)
lin,x., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
Feldlyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 116205)
AUTHORS Lin,X.
TITILE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Dec 17, 1999 this sequence version replaced gi:4512656.
The sequence and annotation of chromosome 2 were merged from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL
(ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
University of Washington), Genscan (Chris Burge,
http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
(http://www.cbs.dtu.dk/services/NetGene2/), searches of the
complete sequence against a peptide database and plant EST
databases at TIGR, and manual curations based on those analyses.
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by two
or more gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
and Satoshi Tabata for helpful assistance. In addition, we would
like to thank the TIGR Bioinformatics Department, especially Dixin
Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
Peterson, Michael Holmes, and Delwood Richardson for software and
database support.

This work was supported by the National Science Foundation,
Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES

source

misc_feature

mRNA

gene

CDS

misc_feature

mRNA

gene

CDS

repeat_region

repeat_region

repeat_region

mRNA

gene

CDS

repeat_region

mRNA

Location/Qualifiers

1..116205

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="II"

complement(<1..2131)

/note="Sequence from clone F14N22"

<1226..>2344

/gene="At2g42650"

/note="F14N22.8; predicted by genscan"

1226..2344

/gene="At2g42650"

/note="hypothetical protein"

/protein_id="AA018664.1"

/db_xref="GI:6598642"

/translation="MSRVSPTVDDAVKALVKGNEKSRTKPOLLEEDGFYLVVAL
KKIPORNFNTAVRIPLPHPILNTEEDSPCLIIDRPESGLTEDAKKIKSENIP
TKVVKLSKLKSDYGSFESKRKLCDSDYDMFSDRRVIMLPKLGKFFQSKKTPVAID
LKHNMWKEOIERACGAMFEMRTGSCSAIKVAKLSMESDDIVENVTATNGVVDLPS
RWKTIKRSIHLKLSLSLPLYTQVTPYLQKIDPLGVEEVKNGEGLAKSDYDDSSKSV
KTKKKNKGIHEVRVYMDSNVSETLGDDEFDRSVGEDEVADLNASGDKKRRKRMSSSKS
AVSGKPDIVKSKNGQSKKLKIDIDESGGGFKAKTKRR"

complement(2132..>116205)

/note="Sequence from clone F7D19"

complement(join(<2761..2943,3035..3164,3258..3372,
3450..3489,3570..3646,3732..>3954))

/gene="At2g42660"

complement(<2761..>3954)

/gene="At2g42660"

/note="F7D19.34; predicted by genscan"

complement(join(2761..2943,3035..3164,3258..3372,
3450..3489,3570..3646,3732..3954))

/gene="At2g42660"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD21740.1"

/db_xref="GI:4512686"

/translation="MAGRSSPEKCSWTWLDLNREETVEKNNEESSVEDEDOTNNVRO
YIRSNMPLRPDPLHLSEYRAVQRLGPDRAIPKLVLEMNLKGLSIAHVSHLOMY
RSKLEPSSRRPGFGAFMSGQRYLMDMIDSRCIPHSDLRHAYNSKTPSVRLNODAVY
TNLGNFMLRPSSWFSGLCRNDPSTENKTLPLLEIRKKNNEKRVREEVSSVKRLKS
MSGGGIQLPFEGNCROKPTDINTMLSLSLSTSSEP"

3681..3724

/rpt_family="(GA)n"

4087..4204

/rpt_family="(TA)n"

5756..5806

/rpt_family="(GA)n"

complement(join(<7700..7818,7906..8112,8195..8348,
8844..8989,9294..>9396))

/gene="At2g42670"

complement(<7700..>9396)

/gene="At2g42670"

/note="F7D19.33; predicted by genscan"

complement(join(7700..7818,7906..8112,8195..8348,
8844..8989,9294..9396))

/gene="At2g42670"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD21739.1"

/db_xref="GI:4512685"

/translation="MPYFAQRLYNCKASFSSDGPITEDALEKYRNVLEKIKPSDVG
I EODAQLARSRSGPLNERNGSNOSPATKYLIHLHECDSFSGIFCMPSSMIPLHNHPG
MTVLSKLVYGSMHVKSVDWLEPQLTEPEDPSQEARPAKLVKTEMTAOSPVTTLYPKS
GGINHCFAITHCAILDLIAPYSSSEHDRNCTYFRKSRRREDLPGELLEVDGEVTVDTW
LLEFQPPDDFVIRIRIPYRGVIRT"

10075..10153

/rpt_family="(GAA)n"

join(10610..10809,11165..11248,11522..11604,11693..11994)


```

150 lucSTyrValProValIleThrIleGluValGlySerSerLeuGlu 166
    :: |||::: || | |||::: ||| :: |||:::
487 CTACTTATATACCGTACATGAGAAGAAAGCTTATTACTGTCTGACAAG 536
    ||| ::| ||||| ::|||::: : : : : :
167 TrpLeuSerCysGlyIleuArgGlyLeuLeuValTyrProGluValLysAr 183
    ||| ::| ||||| ::|||::: : : : : :
537 TGGACCAAAATGGAGAGTTCGAGTCTTGACTCAT.....CTGCCTAA 580
    ||| ::| ||||| ::|||::: : : : : :
183 gleuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlu 200
    :||| :||| ||||| || | |||::: ||| ::|
581 GCTCACCTTTAGGATCATCATGTACATCTTCTCAGCTCGAGAGTGAGA 630
    :||| :||| ||||| || | |||::: ||| ::|
200 euAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGlu 216
    ::| |||::: ||| ::| ::|
631 ATGTAATGATGCA.....TTGGAACGAGAAATATACAGCT 665
    ::| |||::: ||| ::| ::|
217 MetThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLe 233
    : : : : : : : : : : : : : : : : : :
666 CTTAATCATGGGGTTCGAGCAATGGCAGTCAATATTCCTGGATTGCTTAA 715
    : : : : : : : : : : : : : : : : : :
233 uTyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnA 250
    : : : : : : : : : : : : : : : : : :
716 TCATAGAGACACTCAAGGCTAGGAAAAACACTTGAGCTGCCTTT..CAGT 762
    : : : : : : : : : : : : : : : : : :
250 snIleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGly 266
    : : : : : : : : : : : : : : : : : :
763 CCATAGTGACTGAGCGCAGAAATCAAGGAAGACGAACATTTATCAAAAT 812
    : : : : : : : : : : : : : : : : : :
267 CysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGly 283
    ||||| ||||| |||||::: ||| |||
813 AAGAAAGATATGTTAGACAATCTTCTTAATGTTAAAGATGAAGATGAAA 862
    : : : : : : : : : : : : : : : : : :
283 uArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheG 300
    : ||||| ::| ::| ::| : |||
863 AACTTTGGATGACGGAAGAGATTATGATGTCTTCTGATGTATCTTAATG 912
    : ||||| ::| ::| ::| : |||
300 LysGlnHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeu 316
    : : : : : : : : : : : : : : : : : :
913 CTGGTCATGATCTCTCTGGCCACACCAATATGTGGGCTACCGTTTCTTA 962
    : : : : : : : : : : : : : : : : : :
317 GlyLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerTy 333
    : : : : : : : : : : : : : : : : : :
963 CAAGAACACCCGTGAGGTTCTACAAAGACGAAGGCTGAA.....CA 1003
    : : : : : : : : : : : : : : : : : :
333 sGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleL 350
    : : : : : : : : : : : : : : : : : :
1004 AGAAATGATCTCGAAAAGTAGGCGCTGAAGTCAAAAAGGTTATCTCTAA 1053
    : : : : : : : : : : : : : : : : : :
350 euGlu.....GlnLeuLysTyrIleGlyCysValIleLysGluThrLeu 364
    ||| : : : : : : : : : : : : : : :
1054 AAGAAACCCGGAAAATGAGAGTCTTATCCGACGGTTGTCGATGAGACACTT 1103
    : : : : : : : : : : : : : : : : : :
365 ArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysThrPh 381
    |||::: : : : : : : : : : : :
1104 CGAGTCATATAACATTCTCACTCACTGCTTTTCGAGAGGCAAGAGACTGACGT 1153
    : : : : : : : : : : : : : : : : : :
381 eGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerI 398
    |||::: ||||| ||||| ||||| ||||| : : : : :
1154 CGAAATGAATGGCTATTGATCCCAAAAGGTTGGAAGGTTTGGACATGGT 1203
    : : : : : : : : : : : : : : : : : :
398 leCysAspThrHisAspValAlaGluIlePheThrAsnLysGluGluPhe 414
    : : : : : : : : : : : : : : : : : :
1204 TTAGGGATGTCACATCGACCGCTGAAGTCTTTCAGATCCAGAAATTT 1253
    : : : : : : : : : : : : : : : : : :
415 AsnProAspArgPheSerAlaProHisProGluAspAlaSerArgPheSe 431
    : : : : : : : : : : : : : : : : : :
1254 GATCCTGCTAGATGGGATTAATGCTTTGTACCAAAAGCTGCT.....GC 1297
    : : : : : : : : : : : : : : : : : :
431 rPheIleProPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheA 448
    : : : : : : : : : : : : : : : : : :
1298 GTTCCCTTCCTTTGGTGGTGAAGCCATCTATGCCCCAGGAATGATCTCG 1347
    : : : : : : : : : : : : : : : : : :

```

448 IalysileuLeuLysilepheThrValGluLeuAlaArgHisCysAsp 464
 |||||::: :::::|||||::: ::::: :::::
 1348 CFAAGCTCGAAATTTCAATTTTCTTCATCATTTCTCTCAATATATCAG 1397
 465 TrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrValTy 481
 ::::: |||::: ::::: |||::: :::::
 1398 GTGAACGAGACACCCCGAATGTCTAGTGATGTATCTGCCTCACACCAG 1447
 481 rProValAspAsnLeuProAlaArgPheThrHis 492
 |||:::||||| |||||::: ::::: :::::
 1448 ACCCACTGATAATTGCTGTGGCAAGAATTAGTTAT 1481

seq_name: gbp14:SLU54770

seq_documentation_block:	1395 bp	
LOCUS	SLU54770	mRNA
		PLN
		18-OCT-1996

cds.

VERSION U54770.1 GI:1421740

SOURCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

AUTHORS Bishop, G.J., Harris

encodes the first member of a new cytochrome P450 family

MEDLINE 96266705
PREFERENCE 3 (b)333 1 40 13061

AUTHORS Bishop, G.
TITLE Direct Substitution

JOURNAL
Submitted (11-APR-1996) Gerard Bishop, Sainsbury Laboratory, John
Innes Centre, Colney Lane, Norwich, Norfolk NR4 7TU, UK

LEADONES	Location/Qualifiers
SOURCE	1 1395

```
/organism= Lycopersicon esculentum
/strain="GCR758"
```

gene 1 1395 /db_xref=taxon:4081

CDs /gene- uwall 1.1395

```

/gene= uwall
/codon start=1

```

```

/protein_id="AAB17070.1"

```

```

/translation="MAFFLIF
/aa-ali 0.1424/41

```

QSMIDILGKCNIAAVNGSAHKYMRGALSLISPTMIRDOLPKIDEFMRSHLTNMDN

GFQARKIIVNLLRTLIEERRASKEIOHDMGLYIMNEEATRFKLTDEMIIDLITTIYS

LETSRLATIVNGVLKRTQDMEINGYIIPKGWRIVYTRELNDPRLYPDPYSFNPWR

VEAPNGLRIRVSAH"

ORIGIN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

Quality: 444.00 Length: 467

Percent Similarity: 58.672 Percent Identity: 27.409

alignment_block:

```

18 LeuLeuLeuPheLeuAlaAlaIleIleLysLeuTrpAspLeuTyrCysValSe 34
   |||  ::|||  ::|||:::  |||:::  ::
40 CTATGTAATCTTTTGTACTGCTTTATTAAGATGGAATCAAGTCAGTATAA  89
   :  :::::  |||||
34 rGlyArgSpArgSerCysAlaLeuProLeuProProGlyThrMetGlyP  51
   :  :::::  |||||
90 CCAAAAAAAC.....TTGCCCCCTGTGTAATGGGTT  121
   :  :::::  |||||
51 heProPhePheGlyGluThrLeuGlnMetValLeuGlnArgGlySph  67
   ::|||:::|||||  :::::  ::|||
122 GGCACACTTTTGGTGAACACTAGTGTCTTAACTTGGTCCAGTTTC  171
   :  :::::  ::|||  :::::  |||:::  ::|||
68 LeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP  84
   :::::  ::|||  ::|||  ::|||  ::|||:::  ::|||
172 ATGAAAAACCAAGACAGCCAGATATGGAGTTTTTTAAATCACACATACT  221
   :  :::::  ::|||  ::|||  ::|||  ::|||  ::|||
84 eGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle  101
   :|||  |||||:::  |||  :::::  |||  ||||
222 TGGTGTCCACACATGTTTTCATGTGATTCAGAACTGACAGATATATAC  271
   :|||  |||||:::  |||  :::::  |||  ||||
101 euleuGlyAspArgLeuValSerValHisTrpProAlaSerValArg  117
   ||:::  :::::  ::  ::  ::|||  |||:::
272 TAGTGAATGAAGCGAAGAGACTGGTCCACAGATACCACAGTCTATGATA  321
   ||:::  :::::  ::  ::  ::|||  |||:::
118 ThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisLy  134
   |||||  |||||:::  :::::  :::::  |||:::  ||||
322 GATATTTTAGAAAAATGTAATATATTGCAGCTGTCAATGGTTCAAGCTCACAA  371
   :  ::  :::::  :::::  ::|||  |||
134 sGlnArgLysLysValIleMetArgAlaPheSer.....ArgG  147
   |  ::  :::::  :::::  ::|||  |||
372 GTACATGAGGGGTGCATGTTGTTATCCCTAAATTAAGCCCTACAATGATCAGAG  421
   :  ::  :::::  :::::  ::|||  |||
147 IuaIaLeuGluCysTyrValProValIleThrGluGluValGlySerSer  163
   ::  |||  ::|||  |||  |||  ::  |||
422 ACCCAACTT.....TTGCCTAAAAATTGATGAGTTTATGAGATCCAC  462
   |||  ::|||
164 LeuGluGlnTrpLeuSerCysGlyGluArgGlyLeuValTyrProG  180
   |||  ::|||
463 TTAACCAATTGG.....GATAATAAAGTTATGACATTCAAGAGAA  503
   :  ::|||
180 uValLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysG  197
   :::::  :::::  |||  :::::  ::  |||
504 AACCAATAAGATGGCATTTCTATCATCGTTGAAGCAAAATTGCTGTATTG  553
   :  ::|||
197 Iu.....ProGlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGlu  212
   ||  ::|||  |||
554 AATCTACCTCTTTAGCT.....CAAGAATTCATGTCT  585
   :  ::|||
213 AlaPheGluGluMetThrArgAsnLeuPheSerLeuProIleAspValPr  229
   |||  :::::  ::  ::|||  |||||  :::::  |||
586 GAATTTTTCATCTAGTGTAGGCACCTCTTACACTACATCAATCTTCC  635
   :  ::|||
229 oPheSerGlyLeuTyrArgGlyMetLysAlaArg.....AsnL  242
   |  :::::  ::|||  |||||  ::|||
636 AAACACCAACTATCATCGCGGATTTTCAGGCAAGGAAAAATTATGTGAAC  685
   :  ::|||
242 euIleHisAlaArgIleGluGlnAsnIleArgAlaLysIleCysGlyLeu  258
   ||:::  ::|||  ::|||
686 TATTACGAACACTCATAGAGAGAGA.....  711
   :  ::|||
259 ArgAlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeuI  275
   |||||  |||||:::  |||  |||  |||
712 AGAGCTTCAAGGAATTCAA.....CATGATATGCTTGGTTACCTGAT  755
   :  ::|||
275 eGluHisSerTrpGluArgGlyGluArgLeuAspMetGlnAlaLeuLysG  292
   :::::  ::|||  ::  ::|||  ::  ::
756 GAATGAGGAAGCAACAGATTCAAAATTAACAGATGATGAGATGATTGATT  805
   :  ::|||
292 InSerThrGluLeuLeuPheGlyGlyHisGluThrThrAlaSerAla  308
   |||  ::|||  ::|||  ::|||  ::|||  ::|||
806 TAATTATAACT...ATTTGTACTCTGGATATGAACGTGTTCCACCACT  852
   |||  ::|||  ::|||  ::|||  ::|||  ::|||

```

```

309 AlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuGlnIly 325
    ::      ::      |||||      ::|||      |||||:::
853 TCTATGATGCTGTGGAATATCTTCATGATCATCCAAAAGTTCITGAAGA 902
    ::|||:::| |      ::      ::      |||||:::
325 sValArgGluGluLeuLysSerLysGlyLeuLeuCysLysSerAsnGlnA 342
    ::|||:::| |      ::      ::      |||||:::
903 ACTTAGAAAGAA.....CACATGGCTATTAGAGAAAAAGAAAAACCTG 946
    ::      ::|||      ::      ::      ::      ::
342 sPasnLysLeuAspMetGluIleLeuGluGlnLeuLysTyrIleGlyCys 358
    ::      ::|||      ::      ::      ::      ::
947 AGGATCCCTATCGATTACACAGCATTCACAGTTCACACAGAGCT 996
    ::      ::|||      ::      ::      ::      ::
359 ValIleLysGluThrLeuArgLeuAsnProProValProGlyGlyPheAr 375
    |||||      |||||      |||||      |||      |||      ::||
997 GTGATTTTAGAGACCCTCCAGGTTAGCAACAATAGTAATGGGGTTTGAG 1046
    ::      ::      ::      ::      ::      ::      ::
375 gValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyT 392
    |      ::      ::      ::      ::      ::      ::      |||||
1047 AAAAAACAACTCAGATATGGAATAAATGGGTATATCATCTCTAAAGGAT 1096
    ::      ::      ::      ::      ::      ::      ::
392 rPasnValIleTyrSerIleCysAspThrHisAspValAlaGluIlePhe 408
    ||:::      ::      ::      ::      ::      ::      ::
1097 GGAGATATAGCTATATATACAAAGGAGTGAATTAGCATCAAGACTTTAT 1146
    ::      ::      ::      ::      ::      ::      ::
409 ThrAsnLysGluGluPheAsnProAspArgPheSerAlaProHisProG 425
    ::      ::      ::      ::      ::      ::      ::
1147 CCTGATCCATATTCGTTCAATCCATGAGATGATGATGAAGACCTGGA 1196
    ::      ::      ::      ::      ::      ::      ::
425 uAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArgSerC 442
    |      ::      ::      ::      ::      ::      ::      |||||
1197 ACACCAAAAC.....TCATTTTGGTATTGGAGGTGGTACTAGACAAT 1240
    ::      ::      ::      ::      ::      ::      ::
442 ySValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrValGlu 458
    ||      |||||:::      ::      ::      ::      ::      |||
1241 GTCCTGAAAGGAACCTGGTGTAGCAGAAATTTCCACATTTCTTCATTAC 1290
    ::      ::      ::      ::      ::      ::      ::
459 LeuAlaArgHisCysAspTirpGlnLeuLeuAsnGlyProProThrMetLy 475
    ::      ::      ::      ::      ::      ::      ::
1291 TTCGTAACAAAATACAGATGGGAAGAATAGGTGAGATAAACTGATGAA 1340
    ::      ::      ::      ::      ::      ::      ::
475 s 475
    |
1341 A 1341

seq_name: gb_pl2:AF318501

seq_documentation_block:
LOCUS      AF318501      1740 bp      mRNA      PLN      21-FEB-2001
DEFINITION Arabidopsis thaliana ent-kaurenoic acid hydroxylase (KAO2) mRNA,
            complete cds.
ACCESSION  AF318501
VERSION    AF318501.1  GI:13021855
SOURCE     .
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 1740)
            Helliwell,C.A., Chandler,P.M., Poole,A., Dennis,E.S. and
            Peacock,J.W.
            The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes
            three steps of the gibberellin biosynthesis pathway
            Unpublished
TITLE      2 (bases 1 to 1740)
            Helliwell,C.A., Peacock,J. and Dennis,L.
JOURNAL    Direct Submission
AUTHORS     Submitted (02-NOV-2000) Plant Industry, CSIRO, Clunie Ross Street,
            Canberra, ACT 2601, Australia
JOURNAL    location/Qualifiers
FEATURES
            1..1740
            /organism="Arabidopsis thaliana"
            /db_xref="taxon:3702"

```



```
455 eThrValGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsn..GlyP 471
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1419 TCTTCATCATTTCCCTCCTCAATACCGGGTGGAAAGGAGCAATCCCGGAT 1468
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
471 roProThrMetLysThrSerProThrValTyrProValAlaSpAsnLeuPro 487
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1469 GTCCGGTGATG..TTCTTGCCCTCACAATCGACCCAAAGATAATTGCCCTA 1515
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
488 AlaArgPheThr 491
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1516 GCAAGATCACC 1527
```


OM of: US-09-668-482-4 to: N_Geneseq_0601:* out_format : pfs
Date: Nov 6, 2001 3:57 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framet_p2n.model -DEV=xlp
-q=/cgn2_1/USPTO_spool/US09668482/runat_05112001_133354_10917/app-query.fasta_1.1680
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09668482_@CGN1_1_352 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:
Query: US-09-668-482-4
Query length: 497
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 240.270000

Sequence	Strd Orig	ZScore	Escore	len	! Documentation
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09247			2611.00	5029.57	3.5e-272
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12204			2611.00	5029.57	3.5e-272
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV0752			2562.00	4930.47	1.2e-266
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09252			2437.00	4690.54	2.7e-253
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12205			2437.00	4690.54	2.7e-253
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09251			1738.00	3334.52	9.2e-178
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12203			1738.00	3334.52	9.2e-178
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12216			1712.00	3274.65	2.0e-174
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12214			1365.00	2607.04	3.0e-137
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			478.00	892.73	9.4e-42
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			450.50	841.59	6.6e-39
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			436.50	812.53	2.7e-37
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			433.50	807.00	5.6e-37
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			433.50	806.73	5.8e-37
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			433.50	806.71	5.8e-37
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			412.50	765.87	1.1e-34
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			377.00	668.65	2.8e-29
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV09247			348.50	640.37	1.1e-27
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			348.50	640.37	1.1e-27
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			335.50	615.46	2.6e-26
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			335.50	615.46	2.6e-26
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			333.00	610.33	5.0e-26
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			332.00	608.37	6.5e-26
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			332.00	608.37	6.5e-26
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			327.50	615.32	2.7e-26
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			322.00	589.27	7.5e-25
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV10237			322.00	589.27	7.5e-25
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			321.50	590.19	6.7e-25
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			313.50	573.36	5.8e-24
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			310.00	567.10	1.3e-23
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			308.50	562.92	2.2e-23
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			295.50	537.76	5.5e-22
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV09247			293.50	533.88	9.1e-22
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV09247			291.50	530.49	1.4e-21
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV09247			289.00	525.34	2.7e-21
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV09247			288.50	525.29	2.7e-21
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			287.50	522.76	3.8e-21
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247			287.00	522.67	3.8e-21

/SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV09247	286.00	519.56	5.7e-21	1824
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09247	284.50	518.43	6.6e-21	1567
/SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV09247	278.00	504.57	3.9e-20	1745
/SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT:AAV09247	277.50	505.27	3.6e-20	1512
/SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT:AAV09247	277.50	505.27	3.6e-20	1512

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09247
seq_documentation_block:
ID AAV09247 standard; cDNA; 1494 BP.
XX
AC AAV09247;
XX
DT 07-JUL-1998 (first entry)
XX
DE Human cytochrome P450RAI cDNA sequence.
XX
KW Retinoid regulated gene; cytochrome P450 gene; enzyme;
KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1494
FT FT /*tag= a
FT FT /product= "Cytochrome P450RAI"

W09749832-A2.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00488.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
PI Petkovich PM;
XX
DR WPI; 1998-077193/07.
DR P-PSDB; AAW37734.

Identifying DNA encoding inducible or suppressible cytochrome P450 -
by screening for drugs which reduce the catabolism of retinoic acid,
useful in cancer chemotherapy and the treatment of acne and
psoriasis

Example 4; Pages 56-58; 113pp; English.

This nucleotide sequence encodes the human cytochrome P450RAI. Its
expression is dependent on the presence of retinoic acid (RA). The
retinoid-regulated genes such as the inducible cytochrome P450RAI
gene specifically metabolises a derivative of the RA. The cytochrome
P450 gene in general produces enzymes involved in the oxidative
metabolism of endogenous and exogenous compounds. The cytochrome
P450 nucleotide sequence can be used to induce or suppress the
expression of its protein. P450RAI is highly induced by RA in cell
lines and tissues. This allows for development of a drug screen
using promoters and nucleotide sequences to identify drugs which are
useful for reducing the catabolism of RA.

Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other;

alignment_scores:
Quality: 2611.00 Length: 497
Ratio: 5.254 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-668-482-4 x AAV09247 ..

Align seg 1/1 to: AAV09247 from: 1 to: 1494

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuP 17
1 ATGGGGCTCCCGCGCTGCTGGCCAGTGCCTCTGCACCTCTGCTGCC 50
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
51 GCTGCTGCTCTTCTGCTGGCATCAAGCTCTGGGACCTGTACTGCGTGA 100
34 eerglyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
101 GCGGCGCGGAGACCGCAGTGTGCCCTCCCATTTGCCCCCGGAGCTATGGGC 150
51 PheProPhePheglygluThrLeuGlnMetValLeuGlnArgArgLysPh 67
151 TTCCCTCTTCTTGGGGAACCTTGAGATGGTACTGCAGCGGAGAGAGTT 200
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
201 CCTGCAGATGAAGCGCAGGAATAAGGCTTCATCTACAAGACCATCTGT 250
84 heglyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
251 TCGGGCGGGCCACCGGTACGGGTGATGGGCGGAGACAATGTGCGCGCATC 300
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
301 TTGCTCGGAGACGACCGGCTGTGTGCTCCACTGGCCAGCGTCGGTGGC 350
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
351 CACCATCTTGGGATCTGGCTGCTCTTAACTGCACGACTCTCGCACA 400
134 ysglnArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
401 AGCAGCGCAAGAAGGTGATTAATGCGGGCTTCAGCGCGGAGCAGCTCGAA 450
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGlnGlnTr 167
451 TGCTACGTGCGCGGTGATCACCGAGGAAGTGGGAGCAGCCTGGAGCAGTG 500
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
501 GCTGAGCTGCGGCGGAGCGGCGCTCTGTTCTACCCCGAGGTGAAGCGCC 550
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
551 TCAATGTTCCGAATCGCCATGCGCATCTTACTGGGCTGCGAACCCTCAACTG 600
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGluMe 217
601 GCGGCGGACGCGGACTCCGAGCAGCAGCTTGTGGAGGCTTCGAGGAAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
651 GACCCGCAATCTTCTTCTGCTGCTGCATCGACGTGCCCTTCAGCGGGCTGT 700
234 yzArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
701 ACCGGGGCATGAAGCGCGGAACTTCATTCACGCGCGCATCGAGCAGAAC 750
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
751 ATTGCGGCCAAGATCTGCGGGCTGCGGGCATCCGAGGCGGGCCAGGGCTG 800
267 slyAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
801 CAAGACCGCGCTGCAGCTGTGATCGAGCACTCGTGGGAGAGGGGAGAGC 850
284 rgleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
851 GGCTGGACATGACGACACTAAGCAATCTTCAACCGAAGCTCTCTTTGGA 900

301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
901 GGACACGAAACCACGCGGAGTGCAGCCACATCTCTGATCATTACCTGGG 950
317 yLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerLysG 334
951 GCTCTACCCACATGTTCTCTCAGAAAGTGCAGAGAAGCTGAAGAGTAAG 1000
334 lyleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
1001 GTTACTTTGCCAAGACCAATCAAGACAACAAGTTGGACATGAAATTTTG 1050
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
1051 GAACAACCTTAATACATCGGCTGTGTATTGAAGAGACCCTTGACTGAA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
1101 TCCCCCAGTTCCAGAGAGGTTTGGGTTGCTCTGAAGACTTTTGAATTAA 1150
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
1151 ATGATATCAGAGATTCACCAAGGCGTGAATGTATATCTACAGATATCTGAT 1200
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
1201 ACTCATGATGTGGCAGAGATCTTCACCAACAAGAAGAAATTAACTCTGA 1250
417 parPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
1251 CCGATTACGTGCTCTCTCACCCAGAGATGCATCCAGTTCAGCTTCATTC 1300
434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1301 CATTTGGAGAGAGGCTTAGGAGCTGTGTAGGCCAAAGAATTTCAAAATT 1350
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
1351 CTCTCAAAATATTATTACAGTGGAGCTGGCCAGGCATTTGTACTGGCAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
1401 TCTAAATGACCTCCTACAAATGAACAACAGTCCACCGTGTATCTGTGG 1450
484 sPAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
1451 ACAATCTCCCTGCAAGATTCAACCCATTTCATGGGGAATTC 1491
seq_name: /SIDSI/gcgdata/geneseq/NA1998.DAT:AAV12204
seq_documentation_block:
ID AAV12204 standard; cDNA; 1494 BP.
XX
AC AAV12204;
XX
DT 22-JUN-1998 (first entry)
XX
DE Human retinoid metabolising protein hp450RAI cDNA.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; human; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ss.
XX
OS Homo sapiens.
XX
PN W09749815-A1.
XX
PD 31-DEC-1997.
XX

PF 23-JUN-1997; 97WO-CA00440.
XX 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI; 1998-077178/07.
DR P-PSDB; AAW44160.
XX
XX Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 15; Page 55-57; 110pp; English.
XX
CC This cDNA clone codes for a novel human retinoid metabolising
CC protein (see AAW44160) designated hP450RAI. The encoded protein is a
CC retinoid oxidase that has the ability to hydroxylate retinoic acid
CC at the 4 position of the beta-ionone ring and is inducible in
CC epithelial cells exposed to retinoic acid. To isolate the clone,
CC zebrafish P450RAI was used to search an expressed sequence tag
CC database, and an isolated clone was used to screen a cDNA library
CC generated from an NT2 cell line treated with retinoic acid. The
CC hP450RAI gene has been localised to 10q23-24. Zebrafish, human
CC and mouse P450RAI cDNA clones (see AAV12203-05) are claimed. They
CC can be used in the recombinant production of P450RAI. Antisense
CC nucleic acids can be used in a claimed method for inhibiting
CC retinoic acid hydroxylation for the treatment of cancer, actinic
CC keratosis, oral leukoplakia, secondary tumours of the head and/or
CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis.
CC Promoter sequences (see AAV12206-08) for the P450RAI genes are also
CC claimed.
XX
SQ Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other;

alignment_scores:
Quality: 2611.00 Length: 497
Ratio: 5.254 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-668-482-4 x AAV12204 ..

Align seg 1/1 to: AAV12204 from: 1 to: 1494

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
1 ATGGGGCTCCCGGGCTGCTGGCCAGTGGGCTCTGCACCTTCGTGCTGCC 50
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTTrpAspLeuTYrCysValS 34
|||||
51 GCTGCTGCTCTTCCTGCTGCATCAAGCTCTGGGACCTGTACTGCGTGA 100
34 erGlyArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
101 GCGGCGCGACCGCAGATTGGCCCTCCCATTTGCCCGCGGAGACTATGGGC 150
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
151 TTCCCTTCTTTGGGAAACCTTGCAGATGTACTGCAGCGAGGAAGTT 200
67 eleuGlnMetLysArgArgLysTYrGlyPheIleTYrLysThrHisLeuP 84
|||||
201 CCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAGACGATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||

251 TCGGGCGGGCCACCGTACGGGTGATGGCGCGGACATGTGCGGCGCATC 300
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
301 TTGCTCGAGACGACCGGCTGGTGTCCGCTCCACTGGCCAGCGTCGGTGGC 350
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
351 CACCATTTCTGGATCTGGCTGCCCTCTCTAACCCTGCACGACTCTCCGACA 400
134 ySglnArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
401 AGCAGCGCAAGAGGTGATTATCGGGCTTCAGCCGCGAGCACTCGAA 450
151 CysTYrValProValIleThrGlnGluValGlySerSerLeuGlnTr 167
|||||
451 TGCTACGTGCGGTGATCACCGAGGAAGTGGGCAGACCTGGAGCAGTG 500
167 pleuSerCysGlyGluArgGlyLeuLeuValTYrProGluValLysArgL 184
|||||
501 GCTGAGCTGCGCGGAGCGGCTCCTGTCTACCCCGAGGTGAAGCGCC 550
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
|||||
551 TCATGTTCCGAATCGCCTATCGCATCTACTGGGCTGCGAAGCCCAACTG 600
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluJme 217
|||||
601 GCGGGCGACGGGGACTCCGAGCAGCAGCTGTGTGAGAGGCTTCGAGGAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
651 GACCCGCAATCTTCTCTCGCTGCCCATCGACGTGCCCTTCAGCGGGCTGT 700
234 yTrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
701 ACCGGGGATGAAGGCGCGGAACCTCATTCACGCGCGCATCGACAGAAC 750
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
|||||
751 ATTCCGGCCAAAGATCTGCGGGCTGCGGCATCCGAGCGGGCCAGGGCTG 800
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGluA 284
|||||
801 CAAAGACGCGCTGCAGCTGTGATCGAGCACTCGTGAGAGGGGAGAGC 850
284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheGly 300
|||||
851 GGCTGGACATGACAGGCACATAAGCAATCTTCAACCGAATCCTCTTGA 900
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTYrLeuGl 317
|||||
901 GGACACGAAACACGCGCCAGTGCAGCCACATCTGTGATCATTACCTGGG 950
317 yLeuTYrProHisValLeuGlnLysValArgGluLeuLysSerLysG 334
|||||
951 GCTTACCCACATGTTCTCCAGAAAGTGGAGAGAAGCTGAAGAGTAAGG 1000
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
|||||
1001 GTTTACTTTGCAAGACCAATCAAGACAAGTGGACATGGAAATTTTG 1050
351 GluGlnLeuLysTYrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1051 GAACAACCTAAATACATCGGGTGTGTATTAAAGAGACCCCTTCAGTTAA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1101 TCCCCCAGTTCAGAGGGGTTTCGGGTGCTCTGAAGACTTTGAATTAA 1150
384 snGlyTYrGlnIleProLysGlyTrpAsnValIleTYrSerIleCysasp 400
|||||
1151 ATGATATCCAGATTCCCAAGGGCTGGAATGTATATCTACAGTATCTGTGAT 1200

```
401 ThrHisaspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1201 ACTCATGATGTGGCAGAGATCTTCACCAACAAGAAGATTAAATCCTGA 1250
417 PargPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||
1251 CCGATTTCAGTGTCTCCTCACCAGAGATGCATCCAGGTTCACTTCATTC 1300
434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1301 CATTTGGAGGAGGCCCTTAGGAGCTGTGTAGGCCAAAGAAATTTGCAAAAATT 1350
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
|||||
1351 CTTCTCAAAATATTATTACAGTGGAGCTGGCCAGGCATTTGACTGGCAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1401 TCTAAATGGACCTCTCTACAATGMAAACAGTCCACCGGTATCTCTGTGG 1450
484 spAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
|||||
1451 ACAATCTCCCTGCAAGATTCAACCATTTCCATGGGGAATTC 1491
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60752
seq_documentation_block:
ID AAA60752 standard; cDNA; 2124 BP.
XX
AC AAA60752;
XX
DT 31-OCT-2000 (first entry)
XX
DE Human PSEC64 encoding nucleotide sequence SEQ ID NO:1.
XX
KW Human; PSEC64; neuron growth; nerve disease; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 46..696
FT /*tag= a
FT /product= "PSEC64"
FT /note= "protein related to neuron growth"
XX
PN JP2000152790-A.
XX
PD 06-JUN-2000.
XX
PF 19-NOV-1998; 98JP-0329989.
XX
PR 19-NOV-1998; 98JP-0329989.
XX
PA (HERI-) HERIKKUSU KENKYUSHO KK.
XX
DR WPI: 2000-468126/41.
DR P-PSDB; AAB12489.
XX
PT A protein related to the growth of neuron and a gene encoding said
PT protein -
XX
PS Claim 2; Page 8-10; 13pp; Japanese.
XX
CC The present sequence encodes a human protein, designated PSEC64, which
CC is related to neuron growth. The PSEC64 protein and its gene can be used
CC for the development of a preventive agent for use in the treatment of
CC diseases in which nerves are involved.
XX
SQ Sequence 2124 BP; 509 A; 510 C; 547 G; 558 T; 0 other;
```

```
Quality: 2562.00 Length: 497
Ratio: 5.186 Gaps: 0
Percent Similarity: 99.396 Percent Identity: 99.195
alignment_block:
US-09-668-482-4 x AAA60752 ..
Align seg 1/1 to: AAA60752 from: 1 to: 2124
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
46 ATGGGCTCCCGCGCGCTGCTGGCCAGTCGCGCTTCACACTTCGTGCTGCC 95
17 oLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
96 GCTGCTGCTCTTCTCTGGCTGCGATCAAGCTCTGGGACCTGTACTGCGTGA 145
34 ergIArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
146 GCGGCGCGGACCGCAGTGTGCCCCCTCCCATGCCCCCGGAGCTATGGCC 195
51 pheProPhepheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
196 TTCCCTCTCTTTGGGGAACCTTGCAGATGGTACTGCAGCGGAGAGATT 245
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
246 CCTGCAGATGAAGCGCAGGAATATCGGCTTCTACAGAGCCGATCTGT 295
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
296 TCGGGCGGCCACCGGTACGGGTGATGGCGCGGACAATGTGCGGCGCATC 345
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
346 TTGCTGGAGAGCACCGCGGTGGTCCGCTCCACTGGCCAGCGTGGTGGC 395
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
396 CACCATTTGGGATCTGGCTGCTCTTAACCTGCACGACTCCTCCGACACA 445
134 ySGlnArgLysLysValIleMetArgAlaPheSerArgLysAlaLeuGlu 150
|||||
446 AGCAGCGCAAGAAGGTGATATGCGGGCTTCAGCCGCGAGGCACTCGAA 495
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
|||||
496 TGCTACGTGCCGCTGATCACCGAGAAGTGGCAGACGCTGGAGCAGTG 545
167 pLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
546 GCTGAGCTGCGGCGAGCGCGCTCCTGCTACCCCGAGGTGAAGCGCC 595
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
|||||
596 TCATGTTCCGAATCGCCATGCGCATCTACTGGGCTGCGAACCCCAACTG 645
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGluMe 217
|||||
646 GCGGGCGACGGGAGCTCCGAGCAGCAGCTTGTGGAGGCTTTCGAGGA.AT 694
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
695 GACCCGCAATCTCTTCTGCTGCCCATCGACGTGCCCTTCAGCGGGCTGT 744
234 yTrArgIleMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
745 ACCGGGGCATGAAGGCGCGAACCCTCATTCACGCGCGCATCGAGCAGAAC 794
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
|||||
795 ATTGCGGCCCAAGATCTGGGCGCTGCGGCGATCCGAGCGGGCCAGGGCTG 844
```

alignment_scores:

4

4

84 heglyargProthrValargValmetGlyAlaaspAsnValargArgIle 100
|||||
275 TTGGGGCGCCACGCGTGGGTGATGGGGCGGATATATGCGGGCGCATC 324
101 LeuleuglyAspaspArgleuValSerValHistpProAlaSerValar 117
|||||
325 TTGCTGGAGAGACACCGGTGTGTGTGCTGCACTGGCCCGCTGGTGGC 374
117 gThrIleleuglySerGlyCysleuSerAsnleuHisaspSerSerHisL 134
|||||
375 CACCATCCTGGGCGCTGGCTGCTCCCAACTGCACGATTCCTCGCACA 424
134 ysglnArglyslysValIleMetArgAlaphSerArggluAlaleuGlu 150
|||||
425 AGCAGCGAAGAAGTGATATGACAGGCTTCAGCCGCGAGGCACCTCCAG 474
151 CysTyxValProValIleThrGluGluValGlySerSerleuGluInTr 167
|||||
475 TGCTAGCTGCTCGTGATCGCTGAGGAGTCAAGAGTGTGTGGAGCAGTG 524
167 pleuSerCysgllygluArgGlyleuLeuValTyrProgluValLysArgL 184
|||||
525 GCTAAGCTGCGGCGAGCGCGGCTCTGTCTACCCCGAGGTGAAGCGCC 574
184 eumetPheArgIleAlaMetArgIleleuLeuGlyCysgluProgluInleu 200
|||||
575 TCATGTCTCCGCATCGCCATCGCATCTGCTGGCTGGAGCCGGGTCCA 624
201 AlaglyAspGlyaspSerGluGluInleuValGluAlaphGluGluMe 217
|||||
625 GCGGGCGGCGGGAGGAGCAGCAACAGCTCGTGAGGCTTTCGAGAGAGAT 674
217 tThrArgAsnleuPheSerleuProIleaspValProPheSerGlyleuT 234
|||||
675 GACCCGCAATCTTCTTCTTCCCATTTGACGTGCCCTTTAGCGGCTGT 724
234 yArgGlyMetLysAlaArgAsnleuIleHisAlaArgIleGluGluAsn 250
|||||
725 ACCGGGGCGTGAAGCGCGGAACCTTATACACGCGCATCGAGAGAAG 774
251 IleArgAlaLysIleCysGlyleuArgAlaSerGluAlaGlyGluGlyCy 267
|||||
775 ATTGGCGCCAAAGATCCGCGGCTTCAGGCTACAGAGCCGGAATGGGGTTG 824
267 slyAspAlaLeuGlnleuLeuIleGluHisSerTyrGluArgGlyGlu 284
|||||
825 CAAGGACGGCGTGCAGCTCTGATTGAGCACTCGTGGAGAGGGAGAGA 874
284 rgleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheGly 300
|||||
875 GGCTGGATATGCAGGCACTAAACAATCGTCAACAGAGCTCTCTTTGGT 924
301 GlyHisGluThrThrAlaSerAlaAlaThrSerleuIleThrTyrLeuG 317
|||||
925 GGTCATGAACACTACAGCCAGTGTGCGACATCACTGATCACTTACCTAGG 974
317 yLeuTyrProHisValleuGlnLysValArgGluGluLeuLysSerLysG 334
|||||
975 ACTCTACCCACATGTCTCTCCAGAAAGTTCGAGAAGAGATAAAGACAAGG 1024
334 lyLeuLeuCysLysSerAsnGlnAspAsnLysleuAspMetGluIleleu 350
|||||
1025 GCTTACTTTGCAAGAGCAATCAAGACAACAAGTTAGACATGGAAAACTTTG 1074
351 GluGlnleuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1075 GAACAGCTTAATATACATTGGGTGTGTATTAAAGGAGACCTGCGATTGAA 1124
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1125 TCCTCCGGTTCCAGAGGGTTTCGGGTTGCTGTGAGAGACTTTTGAGCTGA 1174
384 snGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSerIleCysAsp 400
XX

|||||
1175 ATGGATACCAGATCCCCAAGGGCTGGAATGTTATTACAGTATCTGTGAC 1224
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1225 ACCCAGCATGTGCGAGATATCTTCACTAACAAGAGGAATTTAATCCGA 1274
417 pArgPheSerAlaProHisProgluAspAlaSerArgPheSerPheIleP 434
|||||
1275 CCGCTTATAGTGCCTCATPCAGAGGATGCTCCCGGTTCAAGTTCATTC 1324
434 ropheGlyGlyleuArgSerCysValGlyLysGluPheAlaTysIle 450
|||||
1325 CATTTGAGAGGCGCTTCGGAGCTGTGTAGCCAAAAGATTGCAAAATT 1374
451 leuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTyrGluLe 467
|||||
1375 CTTCTTAAGATATTACAGTGGAGCTGGCTAGGCACGTGATTTGGCAGCT 1424
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1425 TCTAATGTGACCTCCTTACATGAAGACAGCCCCACTGTGTACCCGTGG 1474
484 sPasnleuProAlaArgPheThrHisPheHisGlyGluIle 497
|||||
1475 ACAATCTCCCTGCCAAGATTCAACCTACTTCCAGGAGATATC 1515
seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12205
seq_documentation_block:
ID AAV12205 standard; cDNA; 1725 BP.
XX
AC AAV12205;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolising protein mp450RAI cDNA.
XX
KW Retinoid metabolising protein; p450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ss.
XX
OS Mus musculus.
XX
FH Key location/Qualifiers
FT CDS 25..1518
FT /*tag= a
XX
PN WO9749815-A1.
XX
PD 31-DEC-1997.
XX
PE 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI; 1998-077178/07.
DR P-PSDB; AAW44161.
XX
PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 15; Page 62-64; 110pp; English.
XX

CC This cDNA clone codes for a novel mouse retinoid metabolising
CC protein (see AAW44160) designated mp450RAI. The encoded protein is a
CC retinoid oxidase that has the ability to hydroxylate retinoic acid
CC at the 4 position of the beta-ionone ring and is inducible in
CC epithelial cells exposed to retinoic acid. The clone was isolated
CC from a retinoic acid-treated P19 teratocarcinoma cDNA library
CC using human P450RAI cDNA (see AAV12204) as probe. Zebrafish, human
CC and mouse P450RAI cDNA clones (see AAV12203-05) are claimed. They
CC can be used in the recombinant production of P450RAI proteins.
CC Antisense nucleic acids can be used in a claimed method for
CC inhibiting retinoic acid hydroxylation for the treatment of cancer,
CC actinic keratosis, oral leukoplakia, secondary tumours of the head
CC and/or neck, non-small cell lung carcinomas, basal cell carcinomas,
CC acute promyelocytic leukaemia, skin cancer, and premalignancy
CC associated with actinic keratosis, acne, psoriasis and/or
CC ichthyosis. Promoter sequences (see AAV12206-08) for the P450RAI
CC genes are also claimed.

XX
SQ Sequence 1725 BP; 408 A; 425 C; 475 G; 417 T; 0 other;

alignment_scores:
Quality: 2437.00 Length: 497
Ratio: 5.035 Gaps: 0
Percent Similarity: 97.384 Percent Identity: 93.360

alignment_block:
US-09-668-482-4 x AAV12205 ..

Align seg 1/1 to: AAV12205 from: 1 to: 1725

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
25 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCGCTTGCACTTCGTGCTGCC 74
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
75 GCTGCTGCTCTTCTGCGCGCTCAAGCTTGGAACCTGTACTGTGTGA 124
34 erGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
125 GCAGCCGCGATCGACGCGCTGCCCTCCCTTGCCCCCGGTACCATGGGC 174
51 PheProPhePheGlyLurThrLeuGlnMetValLeuGlnArgLysPh 67
|||||
175 TTCCATTCTTTGGGAACATTCAGATGGTGTTCAGCGGAGGAAGTT 224
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
225 TCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAGACGCATCTGT 274
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
275 TTGGGGCGCCACGGTGGGGTGATGGCGCGGATATGTGGCGGCATC 324
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
325 TTGCTGGGAGAGCACCGGTTGGTGTCCGTTGCACTGGCCCGCGTCCG 374
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
375 CACCATCTGGGGCGCTGGCTGCTCTCCAACCTGCACGATTCCTCCACA 424
134 ysgLinarGlyLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
425 AGCAGCGAAGAAAGGTGATTATGAGCCCTTCAGCCGAGGCACTCCAG 474
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGlnGlnTr 167
|||||
475 TGCTACGTGCTGCTGATCGCTGAGGAAGTCAGCAGTTGTTGGAGCAGTG 524
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||

525 GCTAAGCTGCGGCGAGCGGCGCTCCTGCTACCCCGAGGTGAAGCCGC 574
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
|||||
575 TCATGTTCGCATCGCCATCGCATCTCTGCTGGGTGCGAGCCGGGTCCA 624
201 AlagLysAspGlyAspSerGlyGlnGlnLeuValGluAlaPheGluGlu 217
|||||
625 GCGGGCGCGCGGAGGAGCAGCAACAGCTCGTGAGGCTTTCGAGAGAT 674
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
675 GACCCCAATCTCTTCTCTCTTCCATTGACGTGCCCTTTAGCGGCTGT 724
234 yrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
725 ACCGGGCGTGAAGCGCGGAACCTTATACACGCCGCCATCGAGGAGAAC 774
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
|||||
775 ATTCGCGCCAAGATCCGCGGCTTCAGGCTACAGAGCCGGATGGGGTTG 824
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGluA 284
|||||
825 CAAGAGCGGCTGCAGCTCCTGATGAGCACTCGTGAGAGGAGGAGAGA 874
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
875 GGCTGGATATGCAGGCACTAAACAAATCGTCAACAGAGCTCTTGTGGT 924
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGl 317
|||||
925 GGTATGAACTACAGCCAGTGTCCGACATCACTGATCACTTACTTAC 974
317 yLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerLysG 334
|||||
975 ACTCTACCCACATGTCTCTCCAGAAAGTTCCAGAGAAGATTAAGACCA 1024
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
|||||
1025 GCTTACTTTGCAAGAGCAATCAAGACAAAGTTAGACATGGAACCTTG 1074
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1075 GAACAGCTTAATAATCATTTGGTGTGTCAATTAAGAGAACCTGGCATTGA 1124
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1125 TCCTCCGGTTCAGAGAGGTTTCGGTGTGCTCTGAAGACTTTTGAGCTGA 1174
384 sngLytYrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
|||||
1175 ATGGATACCAGATCCCAAGGCTGGAATGTTATTACAGTATCTGTGAC 1224
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1225 ACCCAGCATGTGGCAGATATCTTCACTAACAGAGGAATTTAATCCGA 1274
417 parGheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||
1275 CCGCTTTATAGTGCCTTCATCCAGAGGATGCTTCCCGGTTCAAGTTCATTC 1324
434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1325 CATTTGAGGAGGCGCTTCGAGAGCTGTGTAGGCAAGAGTTTGCAAAAT 1374
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
|||||
1375 CTTCTTAAGATATTACAGTGAAGCTGCTAGGCACTGTGATGGCAGCT 1424
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1425 TCTAAATGACCTCTACATGAAGACAAAGCCCACTGTGTACCTGTGG 1474


```

297 LeuLeuphGegLyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuI 313
|||||
964 CTTCTATTGGAGGTCATGAAACACACCGCCAGCACTGCACCTCATTGT 1013
313 eThrTyLeuGlyLeuTyProHisValLeuGlnLysValArgGluL 330
: |||||
1014 CATGTTTCTGGGTGCAACACAGAAAGTGTCAGAGAGTCAAGAGAGG 1063
330 eulysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAsp 346
::: |||
1064 TTCAGGAGAGAGGTGAATGGGCATGTATACACCTGGAAAGGCTTGAGT 1113
347 MetGluLeuLeuGlnGlnLeuLysTyrIleGlyCysValIleLysGluTh 363
|||||
1114 ATGAGAGCTGTGGACCACTGAAGTACACTGATGTGATTAAGAAGAC 1163
363 rLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysT 380
|||||
1164 TCTTAGAATCAACCCCTCTGTCCCGAGAGATTCAAGATCGCACTCAAAA 1213
380 hrPheGluLeuAsnGlyTyGlnIleProLysGlyTyrAsnValIleTy 396
|||||
1214 CTTTGAATTGAATGGTTACCAAAATTCCTAAAGGATGCAACGTCTATTAC 1263
397 SerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysGluI 413
|||||
1264 AGCATCTGTGACACGCACGATGTGGCCGACGCTTTCCAAACAAAGAGGA 1313
413 uPheAsnProAspArgPheSerAlaProHisProGluAspAlaSerArg 430
|||||
1314 GTTCCAGCCGGAGAGATTCATGAGCAAAAGTCTGGAGAGCGGTCCAGGT 1363
430 heSerPheIleProPheGlyGlyLeuArgSerCysValGlyLysGlu 446
|||||
1364 TTACTACATCCCTTCGGAGAGAGATCCAGGATGTGTGGGCAAGAG 1413
447 PheAlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCy 463
|||||
1414 TTCCGCAAGGTGTACTCAAGATCTTTTAACTGAGTTAAACGACAGATTG 1463
463 sAspTyrGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrv 480
|||||
1464 CAATTGGATTTCTCTCAACAGGAGCCCGACCAATGAAAAACAGGCCGACTA 1513
480 aLTyrProValAspAsnLeuProAlaArgPheThrHisPhe 493
::: |||
1514 TTTACCCAGTGGACAATCTCCCTACCAAAATTCACCTAGTTAT 1554

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12203
seq_documentation_block:
ID AAV12203 standard; cDNA; 1850 BP.
XX
AC AAV12203;
XX
DT 22-JUN-1998 (first entry)
XX
DE Zebrafish retinoid metabolising protein zP450RAI cDNA.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; zebrafish; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; ss.
XX
OS Danio rerio.
XX
FH Key Location/Qualifiers
FT CDS 88..1566
FT /*tag= a
XX
PN WC09749815-A1.

```

XX	31-DEC-1997.
PD	
XX	
PF	23-JUN-1997; 97WO-CA00440.
XX	
PR	01-OCT-1996; 96US-0724466.
PR	21-JUN-1996; 96US-0667546.
XX	
XX	
PA	(TOOH) UNIV QUEENS KINGSTON.
XX	
PI	Beckett BR, Jones G, Petkovich PM, White JA;
XX	
DR	WPI; 1998-071178/07.
DR	P-PSDB; AAW12203.
XX	
XX	Retinoid metabolising protein - useful to develop products to treat,
PT	e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT	ichthyosis
XX	
PS	Claim 15; page 52-54; 110pp; English.

This cDNA clone codes for a novel zebrafish retinoid metabolising protein (see AAW44159) designated zp450RAI. The encoded protein is a retinoid oxidase that has the ability to hydroxylate retinoic acid at the 4 position of the beta-ionone ring and is inducible in epithelial cells exposed to retinoic acid. To isolate the clone, transcripts present in fin tissue regenerating in the presence or absence of retinoic acid were compared using a differential display PCR technique, and an isolated clone (see AAV12213) was used to screen a D. rerio 6-18 hr embryo cDNA library. Human (see AAV12204) and mouse (see AAV12205) P450RAI clones have also been identified. The isolated clones can be used in the recombinant production of P450RAI proteins. Antisense nucleic acids can be used in a claimed method for inhibiting retinoic acid hydroxylation for the treatment of cancer, actinic keratosis, oral leukoplakia, secondary tumours of the head and/or neck, non-small cell lung carcinomas, basal cell carcinomas, acute promyelocytic leukaemia, skin cancer, and premalignancy associated with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter sequences (see AAV12206-08) are also claimed.

XX	Sequence	1850 BP	519 A	391 C	454 G	486 T	0 other
50							

```
alignment_scores:
    quality: 1738.00
    ratio: 4.080
    percent similarity: 85.714
    length: 497
    gaps: 5
    percent identity: 68.008
```

```
alignment_block:
```

US-09-668-482-4 X AAV12203

Align seg 1/1 to: AAV12203 from: 1 to: 1850

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
 ||||| :|||:|||||:|||||:|||||:
 88 ATGGGGCTGTACACCCCTTATGGTCACCTTCTCTGCACCACTGTCTACC 137
 17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrcysValS 34
 :||:|||||:|||||:|||||:|||||:||:
 138 CGTTTACTCTTCTCGCCGGGTGAAGTTGTGGAGATGTTAATGATCC 187
 34 eArgLyArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
 ||| :|||:|||||:|||||:|||||:
 188 GACGAGTCGATCCGAACTGCAGAGTCCCTTACCGCCAGGTACCATGGGC 237
 51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
 :||:|||||:|||||:|||||:|||||:|||||:
 238 TTGCCGTTCAATTGGAGAAACGCTCCAGCTGATCCTCCAGAGAAGAAATT 287
 67 eLeuGlnMetLysArgArgLysTyrcylPheIleTyrlsThrHisLeuP 84
 |||||:|||||:|||||:|||||:|||||:
 288 TCTGCCATGAAACGGCAGAAATACGGGTGCATCTTACAAGACGCACCTCT 337

84	heglYarGProthrValarValmetGlyAlaaspasnValarGltle	100
338	TCGGGAACCCGACTGTCAGGGTGATGGAGCTGATATATGAGGCAGATT	387
101	leuLeuGlyaspaspargleuValserValHistpProAlaSerValar	117
388	CTGCTGGGCGAACACAGCTGTGTCTGTTCAGTGGCCAGCATCAGTGAG	437
117	gThrIleuGlysercGlyCysleuSerAsnLeuHisaspSerSerHistl	134
438	AACCATCTGGGCTGTGACACCCTCTCCAATGTCCATGAGCTTCAACACA	487
134	ysGlnArgLysLysValIleMetArgAlapheserArgGluAlaLeuGlu	150
488	AAACAAGAAGAAAGGCCATTATGAGGGCGTCTCTCGACATGCTCTGGAG	537
151	CysTyrValProValIleThrGluGluValGlySerSerleuGluIntr	167
538	CACtACATTCCTCCGTGATCCAGCAGAGGTGAAGAGGCCCATACAGGAATG	587
167	pleu.....SerCysGlyGluArgGlyLeuLeuValTyrProGluV	181
588	GCTGCAAAAAGACTCTCTGC.....GTGCTGTATTATCCAGAAA	625
181	alLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu	197
626	TGAAGAACtCATGTTCGGATAGCTATGAGAACTCTGCTGTTTGA	675
198	Pro..GlnLeuAlaGlyaspGlyaspSerGluGlnLeuValGluAl	213
676	CCAGAGCAAAATAAGACGGAC.....GAGCAAGAATGTTGAAGC	716
213	apheGluGluMetThrArgAsnLeuPheSerleuProIleaspValProp	230
717	TTTTGAGAAATGATCAAAAACCTGTTCTCTTGCCAAATCGACGTTCTT	766
230	heserGlyLeuTyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArg	246
767	TCAGTGTCTGTACAGGGGTTTGAGGGCCAGCAATTTCATTCACTCCAAA	816
247	IleGluGlnAsnIleearGAlaLysIleCysGlyLeuArgAlaSerGluAl	263
817	ATTGAGCAAAACATCAGGAAGAAATTCAGAT...GACGACATGAAAA	863
263	aglYcInGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTyrp	280
864	CGAACAGAAATACAAAGACGCCCTTCAGCTGTGATCGAGAACAGCAGAA	913
280	luArgGlyLuarGleuAspmetGlnAlaLeuLysGlnSerSerThrGlu	296
914	GAGTGCAGCAACCTTTAGTTTGCAGGCGGATGAAGAAGCAGCTACAGAG	963
297	leuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuI	313
964	CTTCTATTGGAGGTCATGAACACCGCCAGCAGCACTGCAACCTCACCTGT	1013
313	eThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluGluL	330
1014	CATGTTTCTGGGTCTGAACACAGAAAGTGTGCAGAAAGTCCAGAGAGAGG	1063
330	eulysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuasp	346
1064	TTCAGGAGAAAGTTGAAATGGCGATGTATACACCTGGAAGGGCTTGAGT	1113
347	MetGluIleleuGluInleuLysTyrIleGlyCysValIleLysGluTh	363
1114	ATGAGAGCTGTTGACCAAGCTGAAGTACACTGGATGTGTGATTAAAGAGAC	1163
363	rLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysT	380
1164	TCTTAGAATCAACCCCTCTGTTCCTCCGGAGGATTTCAGAGTCCGCACTCAAAA	1213

```

380 hrphegluleuAunglyTyrGlnileProlysglyTrpasnValiletyr 396
|||||
1214 CCTTGAATTGAATGGTTACCAAAATTCCTTAAGAGATGGAACGTCAATTAC 1263
397 SerlleCysasptThrhIsaspValalaglulePheThrAsnlysgluGl 413
|||||
1264 AGCATCTGTGACACGCACGATGTGGCCGACGCTTTCCAACAAGAAGAGA 1313
413 upheaSnProcasparGpheSerAlaProHisProgluasPalaserarGp 430
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
1314 GTTCACAGCGAGAGATTCAATGAGCAAAGGCTGTGAGGACGGCTCCAGGT 1363
430 heSerPheilleProPheglylglyLeuarGserCysValglylsclu 446
||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
1364 TTAACATAATCCCCCTTCGAGAGAGATCCAGGATGTGTGTGGCAAAAGAG 1413
447 PheAlalySleleuleulysilePheThrValgluleuAlaArgHiscy 463
|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
1414 TTGCGCCAAAGTGTTACTCAAGATCTTTTAGTTGAGTTAACGCCAGCATTG 1463
463 sAspTrpGlnleuleuAunglyProProthrMetlysthrSerProthrV 480
||:::||| ||| |||||
1464 CAATTCGATTCCTCTCAAACCGACCCCGACACATGAATAACAGGCCGACTA 1513
480 alTyPrOvalasPasnleuProAlaArgPheThrHisPhe 493
:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
1514 TTATACCAGTGGACATCTCCCTACCAAAATTCACCTAGTTAT 1554

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV12216

seq_documentation_block:
ID AAV12216 standard; DNA; 4164 BP.
XX
AC AAV12216;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolising protein mp450RAI genomic DNA.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
XX
XX Mus musculus.
OS
XX WO9749815-A1.
PN
XX 31-DEC-1997.
PD
XX 23-JUN-1997; 97WO-CA00440.
PF
XX 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
XX (TOOH ) UNIV QUEENS KINGSTON.
PA
XX Beckett BR, Jones G, Petkovich PM, White JA;
PI WPI; 1998-077178/07.
DR
XX
XX This sequence comprises a genomic clone of a novel mouse retinoid
CC metabolising protein designated mp450RAI. A cDNA clone (see
CC AAV12205) that includes a coding sequence for the full-length
MP450RAI protein (see AAW44161) is also provided. The encoded

```


CC protein is a retinoid oxidase that has the ability to hydroxylate
CC retinoic acid at the 4 position of the beta-ionone ring and is
CC inducible in epithelial cells exposed to retinoic acid. Zebrafish,
CC human and mouse P450RAI cDNA clones (see AAV12203-05) are claimed.
CC They can be used in the recombinant production of P450RAI.
CC Antisense sequences can be used in a claimed method for inhibiting
CC retinoic acid hydroxylation for the treatment of cancer, actinic
CC keratosis, oral leukoplakia, secondary tumours of the head and/or
CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter
CC sequences (see AAV12206-08) for the P450RAI genes are also claimed.

XX
SQ Sequence 4164 BP; 997 A; 975 C; 1071 G; 1121 T; 0 other;

alignment_scores: length: 1046
 quality: 1712.00 gaps: 10
 ratio: 4.019 Percent Identity: 38.815
Percent Similarity: 40.727

alignment_block:

US-09-668-482-4 x AAV12216 ..

Align seg 1/1 to: AAV12216 from: 1 to: 4164

40 CysAlaLeuProLeuProGlyThrMetGlyPheProPheGlygl 56
 ||||| ||||| :::
73 TGCGGFTCTCTCTCCACCTCCCTCT..... 99
56 uThrLeuGlnMetValLeuGlnArgArgLysPheLeuGlnMetLysArgA 73
 ||||| ||||| ||||| ||||| ||||| |||||
100CAGCGAGGAAGTTTCTGCAGATGAAGCGCA 130
73 rGlySTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVal 89
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
131 GGAATACGGCTCATCTACAAGACGATCTGTTGGCGGCCACGGTG 180
90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAsp 106
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 CGGTGATGGCGCGGATATGTGCGGCATCTGCTGGGAGAGACACCG 230
106 gLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlySerg 123
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 GTTGGTGTGCGTGCACCTGGCCCGCGTGGTCCGACCATCTGGCGCTG 280
123 lCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys..Lys.. 138
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 GCTGCTCTCCAACTGCAGATTCCTCGCACAAAGCAGAAAGAGGTG 330
138 138
331 AGGTGAGCTGGCAACTCTTGCTGGCTGGCAGGAGACCTCATCTATGGCT 380
138 138
381 TGGTTCAGGCAAAATAGAAATGCGGGGAGGGCTAGTCTATGTGGTG 430
138 138
431 GACCAGAACCTCTCTATCTGAGATCCACTTTAGCTTTTCTGCTAGCAG 480
138 138
481 TGGGTAGTCTGGGGGGGACTGAATTTCTTGAAGGTACTCGAAAG 530
138 138
531 CGAAGGGGGGGGGCTGAGGGAAGTAGAGATTGTAACACTCTCTGCTC 580
139ValIleMetArgAlaPheSerArgGlnAlaLeuGln 150
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
581 CTGGGGGTGCTCAGGTGATTATGACAGGCTTCAGCGGAGGCACTCCA 630

150 uCysTyrValProValIleThrGlnGlnValGlySerSerLeuGlnInt 167
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
631 GTGCTACGTGCCCGCTGATCGCTGAGGAAGTACAGCACTGTCTGAGCAGT 680
167 rPheSerCysGlyGlnArgGlyLeuLeuValTyrProGlnValLysArg 183
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
681 GGCTAAGCTGCGGCGAGCGGGCTCTGCTGCTACCCCGAGGTGAAGCCG 730
184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLe 200
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
731 CTGATGTTCCGCATCGCCATGCGCATCTGCTGGCTGCGAGCCGGGTCC 780
200 uAlaGlyAspGlyAspSerGlnGlnLeuValGlnAlaPheGlnGln 217
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
781 AGCGGGCGCGGGGAGAGACAGACAGCTCGTGAGGCTTTCGAGAGA 830
217 eThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
831 TGACCCCAATCTCTCTCTCTCCCATTTGACGTGCCCTTTAGCGGCTG 880
234 TyrArg..... 235
 |||||
881 TACCGGTAAGGGCGGTTTGGCGAGTCGGAGTAGGGGAACGAAGCTCG 930
236Gly.MetLysAlaArg 240
931 GCATCCGCTCACCGCCACGCTCTCTCCCGCTCAGGGCGGTGAAGCGCGG 980
241 AsnLeuIleHisAlaArgIleGlnAsnIleArgAlaLysIleCysgl 257
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
981 AACCTTATACCGCGCGCATCGAGAGAACAATTCCGCCCAAGATCCGCCG 1030
257 yLeuArgAlaSerGlnAlaGlyGlnGlyCysLysAspAlaLeuGlnLeu 274
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1031 GCTTCAGGCTACAGAGCCGATGGGGTTGCAAGACCGCGCTGCAGCTCC 1080
274 euIleGlnHisSerTrpGlnArgGlyGlnArgLeuAspMetGln..... 288
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
1081 TGATTGAGCACTCGTGGGAGAGAGGAGAGAGGCTGATATGACGTGAGA 1130
288 288
1131 AGCAATTTCAAAGGTGCCAAGGCGGGGAGTGCTGTGACTTTCCAGA 1180
288 288
1181 CACACTTCTGGGGTCTCCAAGCCCTGTCAAGGCCCGACCTACTTCCAA 1230
288 288
1231 GTGGCGGGCGATGCTAGGCTTAGAGCTTTTCAACCTGTGGTGTGACCC 1280
288 288
1281 CTTCAAGGAGCAACAACCTTTCAGAAAGGTCCTTAAGACATCTGC 1330
288 288
1331 ATATCCGATATTACATCAAGAACAATACAGTAGCAAAATTACCGTTAT 1380
288 288
1381 GAAGTAGCAACAAGATAATTATTATGTTGGGGTCAACACACAGAG 1430
288 288
1431 AACCGTATTAAAGGTGGCATTTGCTAGAGAGCTGTGAAGGGGTGGC 1480
288 288
1481 TGAGCAATGGGGAAGATCCCAAAGTTCAAAGGGCAAGGCTCATCTACAA 1530

```
288 ..... 288
1531 GGTAAAGCGAGAGACAGATTAAAGGAGTTTGCGTTTTTGTGTGG 1580
288 ..... 288
1581 TCTTTGACTTTCTATGAACAACGATTTTACCCTTGAAGTCTCCGTG 1630
288 ..... 288
1631 CAATATTTCTCAGGTCAGGCTTTGTAAACAGTCTATAACTGCACATCAGA 1680
289 .....AlaLeuLysGlnSerSerThr 295
1681 TCTGTATAAAGCTTCCGTTTATCCCTTAGGCACATAAACAAATCGTCAACA 1730
296 GluLeuLeuPheGlyGlyHisGluThrAlaSerAlaAlaThrSerLe 312
1731 GAGCTCCTCTTTGGTGGTCATGAAGTACAGCCAGTGTGCGACGTCACCT 1780
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGlu 329
1781 GATCACTTAACCTAGGACTCTACCCACATGTCTCCAGAAAGTTCGAGAAAG 1830
329 LuLeuLysSer..... 332
1831 AGATTAAGAGCAAGGTAGGATGATTTCTAGAGGTTCCCATTTGCGTAGGA 1880
332 ..... 332
1881 CATTCCTCTATTAAACACACACACACACCCCACTGTATATAGTTTGCTC 1930
332 ..... 332
1931 GATACACCCAGTACTATGACAGTGAAGATCTGAGAGCTAGTGGGACTGT 1980
332 ..... 332
1981 GGGGAGAGACTCCACCTCGTAATTAAAAAGCAGTTGTGTACTGG 2030
332 ..... 332
2031 GCTCTCTCTTGGGAGAAATTTGACCCTCTCCTCCTCCTCCTCCTCCT 2080
332 ..... 332
2081 CCTCTCTCCTCCTCACACACACACCATCACACCTTTTATAGAGCAAGG 2130
332 ..... 332
2131 TTCTCCTTTCCCTGACCAAGAACATGAATAATGTGATTAGAGCCAATAGC 2180
332 ..... 332
2181 TGATCAGGGTCGCAGTGTGGTGAGGGCTCAGGGTATGACCCTTATATA 2230
332 ..... 332
2231 CCTGATAAGCAACATTTGTCTGGATAATGGTTTAGGCTGAGGAAAGTGTGG 2280
332 ..... 332
2281 AAAGGAAGGCCATCAGGCCATCAGCTCTTCCCTTTTATCCTCTCCCATC 2330
332 ..... 332
2331 CAGACGCCCTTCAGGTTTAGTTAACAGGTAGTCCGTGGGCTGACTTTT 2380
333 .....Lys. 333
2381 TTTTGGAGTGCCAGGGATCCATCACTCACTTTTATCTGTTCCATA 2430
334 .GlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIle 350
```

```
|||||
2431 GGGCTTACTTTGCAAGACCAATCAAGACAACAGTATGACATGAAACTT 2480
350 euGIuGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeu 366
2481 TGGCACAGCTTAATAATACACTGGGTGTGTCAATTAAAGAGACCCTGCGATTG 2530
367 AsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu.. 382
2531 AATCCTCCGGTTCCAGGAGGGTTTCGGGTGCTCTGAAGACTTTTGAGCT 2580
382 ..... 382
2581 GAATGTGAGTGCACCTCCTGTCCCCACCCCAACCCCTCGTCCACGTCGA 2630
382 ..... 382
2631 CTCTGCTATGCTGTGAGCATCAGCTGCCAGAGCAGTGGCTCAGTGGCC 2680
382 ..... 382
2681 TTGACAGTGTCTGCTCCTCTATGTGTAAGGAAACCAATTGTCTCTCTCT 2730
382 ..... 382
2731 CTTAATGCCATTCATGCTAGTAATGACTTTTGTGTGCAAGCTCAGGG 2780
383 .....LeuAsn..GlyTyrG 387
2781 CCGGATGTCAATTCTTAGATTTTCTTTTAAACAGGATAC 2830
387 InIleProLysGlyTyrPasnValIleTyrSerIleCysAspThrHisasp 403
2831 AGATCCCAAGGGCTGGAATGTATTACAGTATCTGTGACACCAACGAT 2880
404 ValAlaGluIlePheThrAsnLysGluIlePheAsnProAspArgPheSe 420
2881 GTGGCAGATATCTTCACTAAACAAGAGAAATTAATCCGACCGCTTAT 2930
420 rAlaProHisProGluAspAlaSerArgPheSerPheIleProPheGlyG 437
2931 AGTGCCTCATCCAGAGAGATGCTCCGGTTACAGCTTCATTTCCATTTGGAG 2980
437 LyGIlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLeuLys 453
2981 GAGGCCCTTCGGAAGCTGTGTAGGCAAGAGTTGCAAAAATTTCTTTAAG 3030
454 IlePheThrValGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGln 470
3031 ATATTACAGTGGAGCTGGCTAGGCACGTGATTGGCAGCTTCTAAATGG 3080
470 yProProThrMetLysThrSerProThrValTyrProValAspAsnLeuP 487
3081 ACCTCTACAATGAAGACAAGCCCACTGTGTACCTGTGTGACCAATCTCC 3130
487 roAlaArgPheThrHisPheHisGlyGluIle 497
3131 CTGCAAGATTTTACCACCTTCCAGGAGATATC 3162
seq_name: /sids1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12214
seq_documentation_block:
ID AAV12214 standard; DNA; 2677 BP.
XX
AC AAV12214;
XX
DT 22-JUN-1998 (first entry)
XX
DE Human retinoid metabolising protein hp450RAI genomic DNA.
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; human; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
```


224 eubProileaspValPropheserGlyleuTyrArg..... 235
|||||
1776 TGCCCATCGACGTGCCCCCTCAGCGGGCTGTACCGGGTAAGCGCGCAAC 1825
235 235
1826 GGGCTGCGGACTAGGGGGCGGAGCTGGGCGTCTGCTCACCGCCGCG 1875
236Gly.MetLysAlaArgAsnLeuIleHisAlaArgIle 247
1876 CTCTCTGCGCTCAGGGCATGAAGCGCGGAACCTCATTACGCGCGCATC 1925
248 GluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaG 264
|||||
1926 GAGCAGAACATTCGCGCCCAAGATCTGCGGGCTGCGGGCATCCGAGCGGG 1975
264 yGlnGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluA 281
|||||
1976 CCAGGGCTGC AAAAGACGCGCTGAGCTGTTGATCGAGCACTCGTGGAGA 2025
281 rgGlyGluArgIleuAspMetGln..... 288
2026 GCGGAGAGCGGCTGACATGCAAGGTGAGTAGCAGCTTCAGACCAGGCACT 2075
288 288
2076 GCGGAGTTTGGTCCCCCTGGCTTCCAAGCGCTGTCCGCGGCCCCCAA 2125
288 288
2126 AGCGCGCGCCTGGGGCCACGCTTCTGGAGTGGGCGCGGCTCAGACTA 2175
288 288
2176 CAGCTATGGAATCCCCGAAGGAGGCTGAGACACCCGCTCAGAGAGAGCTGC 2225
288 288
2226 GGAAGGGGCTGCGGMGAAACTGGGAGCATCCCCCTAGCCTTAMCAGGTT 2275
288 288
2276 TC AAAAGGAAAGTTGGAATTGCAAAAATGTTAATAAGAACCTTGCAT 2325
288 288
2326 TTTAATAAACTAAGACTTTAACTCAGAGATTCCGGTAGRGCGGGTCG 2375
288 288
2376 TACTCGCCTTACTGCTCCAGCTGAACATAAGGAGCGTGCATTTGTTA 2425
288 288
2426 AAGATATTGCTTCCCTGACTTCTGTACAGCAAAACATTAGCCCTCTA 2475
288 288
2476 GTCTTCCCTCCAGACTCTCAGTTCGATTCTGAGTAATCCTTCTGTCAA 2525
289Ala..Leu 290
2526 CCGCAGGAGACTTGTGGAATGTGGTCTCAGCTTAATTTCTTAGGCACCTA 2575
291 LysGlnSerSerThrGluLeuLeuPheGlyGlyHisGluThrThraLase 307
|||||
2576 AAGCAATCTTCAACCGAAGCTCCTTTGGAGGACGAAACACAGGCGCAG 2625
307 rAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuG 324
|||||
2626 TGCAGCCACATCTGTGATCACTTACCTGGGGCTTACCCACATGTCTCC 2675

324 ln 324
2676 AG 2677
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC39029
seq_documentation_block:
ID AAC39029 standard; DNA; 1685 BP.
AC AAC39029;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 23101.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140635.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148365.
PR	13-AUG-1999;	99US-0148684.

PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149115.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	30-AUG-1999;	99US-0151080.
PR	31-AUG-1999;	99US-0151303.
PR	01-SEP-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0158029.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159568.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

alignment_scores:		
Quality:	478.00	Length: 490
Ratio:	1.615	Gaps: 13
Percent Similarity:	60.408	Percent Identity: 27.551

alignment_block:

Align seg 1/1 to: AAC39029 from: 1 to: 1685

15 ValLeuPro...LeuLeuLeuPheLeuAlaAlaIleLeuLysLeuTrpAspLe 30

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

||||: ||| ::| ||||| ||| |||
1201TCACAAACTCATGCTTGTGTTTGAGAGTGGACACAGCCTTGT 1245
443 ValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrValGluLe 459
||| ||| ::| ::| ::|
1246 CCTGTAAGGAAGTACGAGGATGTCGAGATCTCGAGCTTCTCCATTACTT 1295
459 uAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrMetLysT 476
::: ::| |||: ::| ::|
1296 TGTTACGAGATACAGATGGAGGAATAGAGGGGATGAATTAAATG.... 1341
476 hrSerProThrValTyrPro 482
|||: |||
1342GTGTTTCCG 1350

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC42516

seq_documentation_block:

ID AAC42516 standard; DNA; 1648 BP.

XX AAC42516;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 35853.

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.

PR	05-AUG-1999;	99US-0147192
PR	05-AUG-1999;	99US-0147260
PR	06-AUG-1999;	99US-0147303
PR	06-AUG-1999;	99US-0147416
PR	09-AUG-1999;	99US-0147493
PR	09-AUG-1999;	99US-0147935
PR	10-AUG-1999;	99US-0148171
PR	11-AUG-1999;	99US-0148319
PR	12-AUG-1999;	99US-0148341
PR	13-AUG-1999;	99US-0148565
PR	13-AUG-1999;	99US-0148684
PR	16-AUG-1999;	99US-0149368
PR	17-AUG-1999;	99US-0149175
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	20-AUG-1999;	99US-0149723
PR	20-AUG-1999;	99US-0149929
PR	23-AUG-1999;	99US-0149902
PR	23-AUG-1999;	99US-0149930
PR	25-AUG-1999;	99US-0150566
PR	26-AUG-1999;	99US-0150884
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154039
PR	20-SEP-1999;	99US-0154779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156458
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	08-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0159293
PR	13-OCT-1999;	99US-0159294
PR	13-OCT-1999;	99US-0159295
PR	14-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159684
PR	18-OCT-1999;	99US-0159568
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989
PR	25-OCT-1999;	99US-0161404
PR	25-OCT-1999;	99US-0161405
PR	25-OCT-1999;	99US-0161406
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161361
PR	28-OCT-1999;	99US-0161920
PR	28-OCT-1999;	99US-0161992
PR	28-OCT-1999;	99US-0161993
PR	29-OCT-1999;	99US-0162142

```

alignment_scores:      Quality: 436.50      Length: 504
                       Ratio: 1.480      Gaps: 13
Percent Similarity: 58.532      Percent Identity: 27.381

alignment_block:
US-09-668-482-4 x AAC42516      ..

Align seg 1/1 to: AAC42516 from: 1 to: 1648

5 AlaLeuLeuAlaSerAlaLeuCysThrPheValLeuProLeuLeuPh 21
   |||:::      ::::|||      ::      ::|||:::
176 GCTATGACCGAAACGGGTTGATTCATGTCGTGCCCTTTGATAATATT 225
21 e.....LeuAlaAlaIleLysLeuTrpAspL 30
   :      ||      :::::
226 GGGATTGTTGTTTGAATGGGTGCTGAAGACAGATGTAATGTTGATT 275
30 eutyrCysValSerGlyArgAspArgSerCysAlaLeuProLeuPro 46
276 ATGTGTCCAAACTTGGTGAGAAAAACACTAT.....CTGCCACCA 316
47 glyThrMetGlyPheProPhePheGlyGluThrLeuGlnMetVal.... 61
   |||      ::|||:::|||      ::|||:::
317 GGTGATTTGGGATGGCCAGTCATAGCCAAACATGTGGTCTTTCTTAGAGC 366
62 .....LeuGlnArgArgLysPheLeuGlnMetLysArgArgLysT 75
367 TTTCAAACATCTGATCTGATCCTGAATCCTTCATCCAAATCCTACATCACCAGGT 416
75 yrgly.....PheIleTyrLysThrHisLeuPheGlyArgProThrVal 89
   |||||      |||||:::|||||:::|||||      |||      |||
417 ATGGCGGTACCGGGATTTATTAAGCACACATGTTGGGTACCATTGTGTA 466
90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAsp 106
   |||      :::::      |||||:::|||||
467 CTAGTAACAACACACAGAGACTGTAGGCGAGTCTTAACAGATGATGATGC 516
106 gleuValSerValHisTrpProAlaSerValArgThrIleLeuGlySerg 123
517 CTTC...CACATAGGTTGGCCAAATCTACCATGAACCTCATTTGGCAGGA 563
123 LysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLysVal 139
   :::      :::::      :::::|||||:::      :::::
564 AGTCCTTTGTTGGTATCTCTCTTTGAAGAACACAGAAGGCTCAGGCGTTTG 613
140 IleMetArgAlaPheSer...ArgGluAlaLeuGluCysTyrValProva 155
614 ACTTCTGCTCCTGTCATATGGCCCTGAAGCTCTCTGTATACATACAGTT 663
155 IleThrGluGluValGlySerSerLeuGluGlnTrpLeuSerCysGlyG 172
   |||      |||      :::::|||||:::|||||      :::      |||
664 CATTGAAGAACTGTTAATACTGATCTAGAAAAATGGTCCAAAATGGGAG 713
172 IuArgGlyLeuLeuValTyrProGluValLysArgLeuMetPheArgIle 188
   ||      ::|||      ::      :::::|||||      |||:::
714 AAATCGAGTTCTTATCTCAT.....TTGCGTAAAGCTTACGTTTAAGTTC 757
189 AlaMetArgIleLeuLeuGlyCysGluProGlnLeuAlaGlyAspGlyAs 205
758 ATTATGTACATATTTCTCAGCAGTAGAGAGTGAGCATGTATGAT.... 802
205 pSerGluGlnLeuValGluAlaPheGluGluMetThrArgAsnLeuP 222
803 .....TCATTGGAACGGGAGATATACCAACCTTAACATATGAGATTTC 842
222 heSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyMetLys 238
   :::::      |||:::|||||      :::      ::|||:::|||||
843 GAGCAATGGGTATTAATCTTCTCGGTTTGCTTATCATAGAGCTTTGAAG 892
239 AlaArgAsnLeuIleHisAlaArgIleGluGlnAsnIleArgAlaLysI 255
   |||||:::      ::      |||      :::      |||:::|||||
893 GCAAGGAAAAAACTTGTAGCTGCCCTTT...CAGTCCATAGTGACTTAACCG 939

```


80 sThrHisLeuPheGlyArgProThrValArgValMetGlyAlaAspAsnV 97
||||| : : : : :
263 GACGCATCTTTTGGTGAACCGACGATTTCTCAGCTGACCCGGAACGA 312
97 aLArgArgIleLeuLeuGlyAspAspArgLeuValSerValHisTrpPro 113
||| : : : : :
313 ACCGGTTGTTCTTCAGAACGAAGGAGCTTTTGAAGTGTCTTATCCT 362
114 AlaSerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAs 130
||||| : : : : :
363 GCTTCCATTTGTAACTTTTGGGAAACACTCTCTGCTTCTTATGAAAG 412
130 pSerSerHisLysGlnArgLysLysValIleMetArgAlaPheSerArg 147
||| ||||| : : : : :
413 TTCTTTGCATTAACGTAATGCACCTCTCCACCATGAGCTTTGCTAATCTT 462
147 luAlaLeuGlyCysTyrValProVal...IleThrGluGluValGlySer 162
: : : : : : : : : : :
463 CAATCATTAAGACCATCTCATGCTTGATATTGACCGGTAGTCCGGTTT 512
163 SerLeuGluGlnTrpLeuSerCysGlyLuarGlyLeuLeuValTyrPr 179
: : ||| : : : : :
513 AATCTGATCTTGGTCTCT... : : : : :
179 oGluValLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyC 196
||| : : : : :
551 AGAAGCCCAAAAGATTAACGTTTGAGCTTAACGGTGAAAGCAGTTGATGAGCT 600
196 ySGluProGlnLeuAlaGlyAspGlyAspSerGluGlnLeuValGlu 212
: : ||| : : : : :
601 TTGATCCA... : : : : :
213 AlaPheGluGluMetThrArgAsnLeuPheSerLeuProIleAspValPr 229
: : : : :
636 GAGTATCTTCTTGTCAATGGAAGGCTTCTCTCTCTCTCTCTCTCTCTC... 683
229 oPheSerGlyLeuTyrArg...GlyMetLysAlaArgAsnLeuIleHisA 245
||||| ||||| : : : : :
684 .TTCTCCACCACCTTACCGCAAGCCATCCAAGCGCGAGG... : : : : :
245 laArgIleGluGlnAsnIleArgAlaLysIleCysGlyLeuArg... 259
: : : : :
723 ..AAGGTGGCGGAGGCGGTTGACGGGTGGTGATGAAAAAGAGAGGAGAG 770
260 AlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeuLeuIle.. 275
: : ||| : : : : :
771 GAGGAAGAAGGAGCGGAGAGAAGAAGATATGCTTGGCGGCTTGCCTTGC 820
276 ...GluHisSerTrpGluArgGlyGluArgLeuAspMetGlnAlaLeuL 291
: : : : :
821 GCGGATGATGGATTTCGCGATGAGAGATTTGTTGACTTC... : : : : :
291 ySGlnSerSerThrGluLeuLeuPheGlyGlyHisGluThrThrAlaSer 307
: : ||||| : : ||| : : : : :
861TTGGTGGCTTACTTGTCCCGGTTATGAAACAACCTCCACG 902
308 AlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuG1 324
||| : : : : :
903 ATCATGACTCTCGCCGTCAAATTTCTCACCGAGACTCTCTTAGCTCTTGC 952
324 nLysValArgGluGluLeuLysSerLysGlyLeuLeuCysLysSerAsnG 341
: : : : :
953 TCAACTCAAGGAGAG...CATGAAAAGATTAGGCAATGAAGAGGTG 996
341 lAspAsnLysLeuAspMetGluIleLeuGlnLeuLysTyrIleGly 357
: : : : :
997 ATTGCTATAGTCTTGAATGAGTGATTACAAGTCAATGCATTCACACAA 1046
358 CysValIleLysGluThrLeuArgLeuAsnProProValProGlyGlyPh 374
||||| : : : : :
1047 TGTGTGGTTAATGAGACGCTACGAGTGGCTAACATCATCGCGGTGTTT 1096
374 eArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysG 391

||||| : : : : :
1097 CAGACGTGCAATGACGGATGTTGAGATCAAGGTTATAAATTCCAAAAG 1146
391 lYTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaGluIle 407
||||| : : : : :
1147 GGTGAAAGTATTCTCATCTGTTTGAAGCGGTTCAITTTAGACCCAAACAC 1196
408 pheThrAsnLysGluGluPheAsnProAspArgPheSerAlaProHisPr 424
||| : : : : :
1197 TTCAAAGATGCTCGCACTTTCAAACCTTGAGATGGCAGAGCAACTCGGT 1246
424 oGluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArgS 441
: : : : :
1247 AACGACAGGCCCTTCTAATGTGTTCACACCGCTTGTGAGGGCCAAGGC 1296
441 erCysValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrVal 457
||| ||| ||| : : : : :
1297 TATGTCCCGGTTACGAGCTGGCTAGGTTGCACTCTGTTCCTTCC 1346
458 GluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrMe 474
: : ||| : : : : :
1347 CGCCTAGTGACAGGCTTCAGTTGG... : : : : :
474 tLysThrSerProThrValTyrProValAspAsnLeuProAlaArgPheT 491
: : : : :
1382 GCAAGACAAGCTGTGTTCTTCTTCCAACTACAGAAGCAGAAACGGTACC 1431
491 hrHisphe 493
|||
1432 CGATCTTC 1439
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC48157
seq_documentation_block:
ID AAC48157 standard; DNA; 1646 BP.
XX
AC AAC48157;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 56467.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 29-JUN-1999; 99US-0141287.
PR 30-JUN-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

alignment_scores:

Quality:	433.50	Length:	486
Ratio:	1.521	Gaps:	13
Percent Similarity:	58.642	Percent Identity:	26.543

alignment_block:

US-09-668-482-4 X AAC48157

Align seg 1/1 to: AAC48157 from: 1 to: 1646

19 LeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValSerGI 35
 |||||:||||:|||||
 77 CTCCTCCCTCTCTCTCCATCGCCGCCGCTTCCTCCTCTACTCCGC.. 124
 35 yArgAspArgSerCysAlaLeuProLeuProGlyThrMetGlyPhe 52
 ||| ||| :||| |||||:|||||
 125 .CGTACACGTTAACCGTCGGATGGGTCTGCTCCGGGAAGCCTGGTCTCC 173
 52 ropPheGlyGluThrLeuGlnMetVal.....LeuGln 63
 ||:|||||:|||||:|||||:|||||
 174 CTCGTATAGAGAGACTTTTCAGCTGATCGGAGCTTACAACAAGAGAAC 223
 64 ArgArgLysPheLeuGlnMetLysArgArgLysTyrGlyPheIleTyr 80
 :||| |||||:||||| :||| ||||| :|||
 224 CCGTGAACCTTTCATCGACGAGAGAGTACGCCGGTACGGTTCCGTTTCAT 273
 80 sThrHisLeuPheGlyArgProThrValArgValMetGlyAlaAspAsn 97
 ||||| |||||:|||||:|||||:|||||
 274 GACGCATCTTTTGGTGAACCGACGATTTTCTCAGCTGACCCGGGAACGA 323
 97 aLArgArgIleLeuLeuGlyAspAspArgLeuValSerValHisTrpPro 113
 ||| :||| :|||:||||| :||| :|||
 324 ACCGGTTTGTCTTCAGAACGAAGGAGCTTTTGAGTGTCTTATCTCT 373
 114 AlaSerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAs 130
 |||||:||||:|||||:||||| :||| :|||
 374 GCTTCCATTTGTAAACCTTTGGGGAACACACTCTCTGCTCTTATGAAGG 423
 130 pSerSerHisLysGlnArgLysLysValIleMetArgAlaPheSerArg 147
 ||| |||||:||||| :||| :|||
 424 TTCTTTGCATAAAGCATATGCACCTCTCTCACCATGAGCTTGCATAATTCTT 473
 147 IuAlaLeuGluCysTyrValProVal..IleThrGluGluValGlySer 162
 :||| :|||:||||| :||| :|||
 474 CAATCATTTAAAGACCATCTCATGCTGATATTGACCGGTGATCCGGTTT 523
 163 SerLeuGluGlnTrpLeuSerCysGlyGluArgGlyLeuLeuValTyrPr 179
 :|||:||||:||||| :||| :|||
 524 AATCTGATTTCTTGGTCTCT.....CGTGTCTCTCCATCGA 561
 179 ogLiuValLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyC 196
 |||||:||||:|||||:|||||:|||||
 562 AGAAGCCAAAGATAACGTTTACCTTAACGGTGAAGCAGATTGATGAGCT 611
 196 ysgLiuProGlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGlu 212
 :||| ||| :|||:||||| :||| :|||
 612 TTGATCCA.....GGGGAATGGAGTGAAGCTTTAAAGGAA 646

213 AlaphegluGlumethrArgAsnLeuphSerleuProIleAspValPr 229
::: ::::: ||||| :::::
647 GAGTATCTTCTTGTCATCGAAGGCTTCTTCTCTCTCCTCCCTC. 694
229 opheSerGlyLeuTYrArg..GlyMetLysAlaArgAsnLeuIleHisA 245
||||| ||||| ::::: ||||| :::
695 .TTCTCCACCACTTACC GCAAAAGCCATCCAAAGCGCGGAGG. 733
245 laArgIleGluGlnAsnIleArgAlaLysIleCysGlyLeuArg..... 259
::: ::::: ::::: |||
734 .AAGGTGCGGAGGCGGCTTGACGGGTGGTGATGAAAAGAGGAGAGAG 781
260 AlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeuIle.. 275
::: ||::: ||||| ||| ||:::
782 GAGGAAGAGAGAGCGGAGAGAAAGAAAGATATGCTTGCGGCTGCTGC 831
276 ...GluHisSerTRpGluArgGlyLysArgLeuAspMetGlnAlaLeu 291
::: ::::: ||| ::||:::
832 GCGGATGATGGAATTTCCGATGAAGAGATGTTGACTTC. 871
291 ysgInserSerThrGluLeuLeupheGlyGlnHisGluThrThraIAser 307
::: ||||| ::||::: ||||| :::::
872TTGGTGCTTACTTGCGCGGTATGAAACAACCTCCACG 913
308 AlaAlaThrSerLeuIleThrTYrLeuGlyLeuTYrProHisValLeuG 324
||| ::::: |||||
914 ATCATGACTCTCGCCGCTCAAAATTTCTCAACCGAGACTCCTTAGCTTC 963
324 nLysValArgGluGluLeuLysSerLysGlyLeuLeuCysLysSerAsnG 341
::: ||||| ::::: |||||
964 TCAACTCAAGGAAGAG....CATGAAGAATTAGGGCAATGAAGAGTG 1007
341 InAspAsnLysLeuAspMetGluIleLeuGlnLeuLysTYrIleGly 357
::: ::||::: ::::: |||||
1008 ATTCTGATAGTCTTGAATGAGAGTGAATTACAAGTCAATGCCATTACACA 1057
358 CysValIleLysGluThrLeuArgLeuAsnProProValProGlyGlyPh 374
||||| ::||::: ||||| ::|||
1058 TGTGTGCTTAATGAGACGCTACGAGTGCTTAACATCATCGCGGTGTTT 1107
374 eArgValAlaLeuLysThrPheGluLeuAsnGlyTYrGlnIleProLysG 391
||||| ::||::: ||||| ::|||
1108 CAGACGTGCAAATGACGGATGTGAGATCAAAAGTTATAAAATTCAAAAG 1157
391 lyTPAsnValIleTYrSerIleCysAspThrHisAspValAlaGluIle 407
||||| ::||::: ||||| ::|||
1158 GGTGCAAGTATTCTCATCGTTTGAGCGGTTCAATTAGACCCAAACAC 1207
408 pheThrAsnLysGluGluPheAsnProAspArgPheSerAlaProHisPr 424
||| ::::: ||||| ||||| :::::
1208 TTCAAAGATGCTCGACACTTCAACCCCTTGAGATGGCAGAGCACTCGGT 1257
424 oGluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArgS 441
::: ||| ||||| ||||| |||||
1258 AACGACAGGCCCTTCTAATGTGTTCACACCCGTTGGTGAGGGCCAAAGC 1307
441 erCysValGlyLysGluPheAlaLysIleLeuLysIlePheThrVal 457
||| ||| ||||| ::||::: |||||
1308 TATGTCCCGGTAGAGCTGGCTAGGGTTGCACCTCTGTGTTTCTTCAC 1357
458 GluLeuAlaArgHisCysAspTRpGlnLeuLeuAsnGlyProProThrMe 474
::||::: ::||| ||:::
1358 GCGCTAGTGACAGGCTTCAGTTGG.....GTTCCCTGCAGA 1392
474 tLysThrSerProThrValTYrProValAspAsnLeuProAlaArgPheT 491
::: ::::: ||||| ::||::: |||||
1393 GCAAGACAAAGCTGTTTCTTCCAACTACAGAAGACGCAAAACGGTACC 1442
491 hrHisPhe 493
|||
1443 CGATCTTC1450

seq_name: /SIPs1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC47986
seq_documentation_block:
ID AAC47986 standard; DNA; 1649 BP.
XX
AC AAC47986;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55845.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.


```
966 TCAACTCAAGGAAGAG.....CATGAAAAGATTAGGGCAATGAAGAGTG 1009
341 lnbspasnLysLeuaspMetGluIleLeuGluInLeuLysTyrIleGly 357
      ::::: |||||::: ::: ::::: :::
1010 ATTCGTATAGTCTTGAAATGAGATGATTACAAATCAATGCCATTCCACACAA 1059
358 CysValIleLysGluThrLeuArgLeuasnProProValProGlyGlyPh 374
      |||||::: ||||| |||||::: ::: ||| ||
1060 TGTGTGGTTAATGAGACGCTACGAGTGGCTTAACATCATCGCGGTGTTT 1109
374 eargValAlaLeuLysThrPheGluLeuasnGlyTyrGlnIleProLysG 391
      ||| |||::: |||::: |||||::: ||||| |||||
1110 CAGACGTGCAATGACGGATGTGAGATCAAAAGTTATAAAATCCAAAAG 1159
391 lYTrpAsnValIleTyrSerIleCysAspThrHisaspValAlaGluIle 407
      |||||::: |||||::: |||||::: |||||
1160 GTTGGAAAGTATTCTCATCGTTTAGACGGGTTCATTAGACCCAAACCAC 1209
408 pheThrAsnLysGluGluPheAsnProAspArgPheSerAlaProHisPr 424
      ||| ::: ::: ||||| |||::: :::
1210 TTCAAGATGCTCGCACTTCAACCCCTTGAGATGGCAGACCAACTCGGT 1259
424 ogIuaspAlaSerArgPheSerPheIleProPheGlyGlyLeuArgS 441
      ::: ||| ||||| ||||| ||||| |||
1260 AACGACAGGCCCTTCTAATGTGTTCACACCGTTTGGTGAGGGCCAAAGGC 1309
441 erCysValGlyLysGluPheAlaLysIleLeuLysIlePheThrVal 457
      ||| ||| |||||::: |||||::: |||||
1310 TATGTCCCGGTTACGAGCTGGCTAGGGTTGCACCTCTGTCTTCTTCAC 1359
458 GluLeuAlaArgHisCysAspTrpGlnLeuLeuasnGlyProProThrm 474
      ::: |||::: ::: ||| |||||
1360 CGCCTAGTGACAGGCTTCAGTTGG.....GTTCCCTGCAGA 1394
474 tlysthrSerProThrValTyrProValaspasnLeuProAlaArgPhe 491
      ::: ::: ::: |||||::: |||||
1395 GCAAGACAAGCTGTTTCTTCCAACTACAGAAGCGCAAAACGGTACC 1444
491 hrHisPhe 493
      |||
1445 CGATCTTC 1452
```


OM of: US-09-668-482-4 to: Issued_Patents_NA:* out_format : pfs
Date: Nov 6, 2001 2:29 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09668482/runat_05112001_133353_10897/app_query.fasta_1.1680
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09668482@cgn1_1_104 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:
Query: US-09-668-482-4
Query length: 497
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 149.780000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation	..	1494	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-5				2611.00	5650.04	0.0	1494	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-5				2611.00	5650.04	0.0	1494	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-31				2437.00	5268.72	8.1e-286	1725	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-724-466B-3				1738.00	3743.60	7.2e-201	1850	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-3				1738.00	3743.60	7.2e-201	1850	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-38				1712.00	3676.12	4.1e-197	4164	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-36				1365.00	2925.35	2.7e-155	2677	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-11				614.00	1314.78	1.4e-65	351	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-11				614.00	1314.78	1.4e-65	351	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-622-166A-1				433.50	900.96	1.6e-42	1608	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-991-677-3				348.50	713.51	4.3e-32	1883	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2				344.00	600.54	8.4e-26	4403765	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1				344.00	600.52	8.5e-26	4411529	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-091-432-1				322.00	656.05	6.8e-29	1838	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2				303.00	511.14	8.0e-21	4403765	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1				303.00	511.12	8.1e-21	4411529	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-3				289.00	584.14	6.9e-25	1831	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-1				276.00	555.75	2.6e-23	1838	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-313-075A-49				264.50	531.27	6.1e-22	1757	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-37				260.00	534.02	4.3e-22	683	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-292-768-67				252.50	505.82	1.6e-20	1665	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-292-768-69				252.50	505.82	1.6e-20	1665	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-882-164D-3				251.50	503.64	2.1e-20	1665	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-292-768-3				251.50	503.64	2.1e-20	1665	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-606-505B-63				251.00	499.00	3.8e-20	2174	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-616-990-63				251.00	499.00	3.8e-20	2174	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-274A-1				244.50	489.32	1.3e-19	1551	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-274A-1				244.50	489.32	1.3e-19	1551	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-091-432-3				244.50	473.34	1.0e-18	5156	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-912-900-29				242.50	483.31	2.9e-19	1755	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-285-309-29				242.50	483.31	2.9e-19	1755	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-502-046-29				242.50	483.31	2.9e-19	1755	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-033-055A-3				240.00	477.82	5.8e-19	1761	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-292-768-63				240.00	477.81	5.8e-19	1761	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-274A-27				238.50	476.23	7.1e-19	1551	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-274A-27				238.50	476.23	7.1e-19	1551	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-313-075A-29				238.50	475.29	8.0e-19	1665	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-9				238.50	474.34	9.0e-19	1788	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-17				237.00	471.71	1.3e-18	1704	!

/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-991-677-1 + 235.00 467.32 2.2e-18 1708
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-881-784-5 + 235.00 466.91 2.3e-18 1762
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-292-768-1 + 235.00 466.91 2.3e-18 1762
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-5 + 234.50 466.74 2.4e-18 1644
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-292-768-65 + 234.00 464.72 3.1e-18 1762

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-5

seq_documentation_block:

Sequence 5, Application US/08724466B
Patent No. 6063606
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
ZIP: M5L 1A9
COUNTRY: Canada

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 1494 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-5

alignment_scores:

Quality: 2611.00 Length: 497
Ratio: 5.254 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-668-482-4 x US-08-724-466B-5 ..

Align seg 1/1 to: US-08-724-466B-5 from: 1 to: 1494

1 MetcGlyLeuProAlaLeuLeuAlaSerAlaLeuGlyThrPheValLeuPr 17
|||||
1 ATGGGGCTCCCGGGCTGCTGGCCAGTGGCGCTCTGCACCTTCTGCTGCC 50
17 OleuLeuLeuPheLeuAlaAlaIleLeuLeuTrpAspLeuTyrCysValS 34
|||||
51 GCTGCTGCTCTTCTGCTGGCTGCATCAAGCTCTGGGACCTGTACTGGCTGA 100
34 erGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
101 GCGGCGCGACGACGAGTGTGCTCCATTCGCCCGGCGGACTATGGGC 150
51 PheProPhePheGlyGluThrLeuGluMetValLeuGlnArgArgLysPh 67
|||||

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-882-164D-5

alignment_scores:
Quality: 2611.00 Length: 497
Ratio: 5.254 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-668-482-4 x US-08-882-164D-5 ..

Align seg 1/1 to: US-08-882-164D-5 from: 1 to: 1494

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
1 ATGGGGCTCCCGCGCTGCTGGCCAGTGCCGCTCTGCACCTTCGCTGCC 50
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
51 GCTGCTGCTCTTCTGCTGCTGCATCAAGCTTGCGACCTGTACTGCGTGA 100
34 erGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
101 GCGGCGCGGACCGCAGTGTGCCCTCCCATTTGCCCGCGGACTATGGGC 150
51 pheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
151 TTCCCTCTCTTGGGGAACCTTGACAGATGGTACTGCACGAGGAAGTT 200
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
201 CCTGCAGATGAAGCGCAGGAATAACGGCTTCATCTACAAGACGATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
251 TCGGGCGGGCCACCCTACGGGTGATGGCGCGGACAAATGTGCGGCGCATC 300
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
301 TTGCTCGGAGACGACCGCGCTGCTCGGTCCACTGCGCCACGCTCGGTGCG 350
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
351 CACCAATTCTGGGATCTGGCTGCTCTTAACCTGCACGACTCTCGCACAA 400
134 ySGlnArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
401 AGCAGCGCAAGAAGGTGATGATGCGGGCCTTCAGCGCGGAGGCACTCGAA 450
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
451 TGCTACGTGCGCGGTGATCACCAGAGGAAGTGGCAGACCTGGAGCAGTG 500
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
501 GCTGAGCTGGGGCGAGCGCGGCTCTGCTACCCCGAGGTGAAGCGCC 550
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
551 TCATGTTCGGAATCGCCATGCGCATCTACTGGGCTGCGAACCCTCAACTG 600
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGluMe 217
601 GCGGGCGACGGGGAAGTCCGAGCAGCAGCTTGTGAGAGCCCTTCGAGAAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
651 GACCCGCAATCTCTTCTGCTGCTGCCATGACGAGTGCCCTTCAGCGGCTGT 700
234 yrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250

|||||
701 ACCGGGGCATGAAGCGCGGAACCTCATTCACGGCGGCATCGAGCAGAAC 750
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
751 ATTGCGGCAAGATCTGGCGGCTGGGGCATCCGAGCGGGCCAGGGCTG 800
267 slyAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
801 CAAAGACGCGCTGCAGCTGTGATCGACACTCGTGAGAGAGGGAGAGC 850
284 rGlyAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
851 GCGTGACATGCAGGCACTAAAGCAATCTTCAACCGAAGCTCTTTTGA 900
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
901 GGACACGAACACCGCGCAGTGCACCCACATCTGTGATCAGTACTGCGG 950
317 yLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerLysG 334
951 GCTTACCACATGTCTTCCAGAAAGTGCAGAGAGGCTGAAGAGTAAGG 1000
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
1001 GTTACTTTGCCAAGACCAATCAAGACAAAGTTGGACATGGAAATTTTG 1050
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
1051 GAACAACCTTAATACATCGGGTGTGTTAAGAGAGACCCCTTCGACTGAA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
1101 TCCCCAGTTCGAGGAGGTTTCGGGTGCTCTGAAGACTTTTGAATTAA 1150
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
1151 ATGGATACCAGATTCCCAAGGGCTGGAATGTTACTACAGTATCTGTGAT 1200
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
1201 ACTCATGATGTGGCAGAGATCTTCACCAACAAGAAAGAAATTTAATCTGA 1250
417 parGheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
1251 CCGATTGAGTGTCTCTCACCCAGAGATGCATCCAGGTTCAAGCTTCATTC 1300
434 ropheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1301 CATTTGAGAGGCGCTTGAAGCTGTGTAGCAAGAATTTGCAAAATTT 1350
451 leuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
1351 CTTCTCAAAATATTACAGTGGAGCTGGCCAGCATTTGTGACTGGCAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
1401 TCTAAATGACCTCTTACAATGAATAACAGTCCACCGTGTATCTCTGTGG 1450
484 spAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
1451 ACAATCTCCCTGCAAGATTCAACCATTTCCATGGGGAATATC 1491
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-31
seq_documentation_block:
; Sequence 31, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43

```

CORRESPONDENCE ADDRESS:
ADDRESSSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/000010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEO ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1725 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-31
```

alignment_scores:
Quality: 2437.00 Length: 497
Ratio: 5.035 Gaps: 0
Percent Similarity: 97.384 Percent Identity: 93.360

alignment_block:
US-09-668-482-4 x US-08-882-164D-31 ..

Align seg 1/1 to: US-08-882-164D-31 from: 1 to: 1725

```

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
25 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCGCTCTGCACCTTCGTGCTGCC 74
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
75 GCTGCTGCTCTTCTTGGGGCTCAAGCTCTGGAGCTGTACTGTGTGA 124
34 erGlyArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
125 GCAGCCGCGATCGCAGCTGGCGCTCCCTTGGCCCCCGGTACCATGGGC 174
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
175 TTCCCATTTCTTGGGGAACATTCAGATGTGTCTCAGCGGAGGAAGTT 224
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
225 TCTGCAGATGAAGCGCAGGAATAACGGCTTCATCTACAGACGCATCTGT 274
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
275 TTGGGGGGCCACGGGTGGGGGTGATGGGGCGGATATGTGGCGCATC 324
101 leuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
325 TTGCTGGAGAGCACCGGTGTGTGTCGTCACACTGGCCCGCGTGGTGGCG 374
```

```

117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
375 CACCATCTGGGGCTGGCTGCCCTCTCCAACTGCACGATTCCTCGCACA 424
134 ySglnArgLysLysValIleMetArgAlaPheSerArgLAlaLeuGlu 150
|||||
425 AGCAGCGAAGAGGTGATTTATGCAGCGCTTCAGCCGCGAGGCACTCCAG 474
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGlnTr 167
|||||
475 TGCTACGTGCTGCTGATCGCTGAGGAAGTCACAGTTGTCTGAGCAGTG 524
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
525 GCTAAGCTGGCGCGAGCGCGGCTCTGCTCTACCCGAGGTGAAGCGCC 574
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
|||||
575 TCATGTTCGCGATCGCCATCGCATCCCTGCTGGGCTGCGAGCCGGTCCA 624
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGlu 217
|||||
625 GCGGGCGGGCGGAGGAGCAGCAGCAGCAGCTGTGAGGCTTTCGAGGAGAT 674
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
675 GACCCGCAATCTCTCTCTCTTCCCATGTACGTGCCCTTTAGCGGCTGT 724
234 yTrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
725 ACCGGGGCGTGAAGCGCGCGGAACCTTATACAGCGCGCATCAGAGAAC 774
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
|||||
775 ATTGCGCCAGATCCCGCGCTTCAGGCTACAGAGCCGATGGGGTTTG 824
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
|||||
825 CAAGGACGCGCTGCAGCTCCTGATGTGACACTCGTGGAGAGGGGAGAGA 874
284 rGluAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
875 GGCTGATATGCAGGCACTAAACAAATCGTCAACAGAGCTCCTTTTGGT 924
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGl 317
|||||
925 GGTATGAATACTACAGCCAGTGTGCGACATCACTGATCACTTAACCTAGG 974
317 yLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerLysG 334
|||||
975 ACTTACCCACATGTCTCTCCAGAAAGTTTCGAGAAGATTAAGCAAGG 1024
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
|||||
1025 GCTTACTTTGCAAGAGCAATCAAGCAACAAGTTAGACATGAAACTTTG 1074
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1075 GAACAGCTTAATATACATTTGGGTGTTCATTTAAGAGAGACCCCTGGATTGAA 1124
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1125 TCCTCGCGTTCCAGAGAGGTTCGGGTGTGCTCTGAAGACTTTTGAAGCTGA 1174
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
|||||
1175 ATGGATACCAAGATCCCAAGGGCTGGAATGTATTATTACAGTATCTGTGAC 1224
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1225 ACCCAGATGTGGCAGATATCTTCACTAACAAGAGGAAATTAAATCCCGA 1274
```

417 parpheserAlaProHisProGluAspAlaSerArgpheserPheIleP 434
||||| :|||||
1275 CCGCTTATAGTCCATCCAGAGGATGTCGCCGTTCCAGCTTCATTC 1324
434 ropheglYglYleuArgSerCysValGlyLysGluPheAlaIle 450
|||||
1325 CATTTGAGAGGCGCTTCGGAGCTGTGTAGCAAAAGATTGCCAAAATT 1374
451 leuLeuYsIlePheThrValGluLeuAlaArgHisCysAspTrpGln 467
|||||
1375 CTCTTAAGATATTACAGTGGAGCTGGCTAGGCACCTGTGATTGGCAGCT 1424
467 uLeuAnGlyProProThrMetLysThrSerProThrValTyrProVal 484
|||||
1425 TCTAATGACCTCCTCAATGAAGACAGCCCACTGTGTACCCCTGTGG 1474
484 spAnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
|||||
1475 ACAATCTCCCTGCAAGATTCACCTACTTCCAGGAGATATC 1515

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-724-466B-3

seq_documentation_block:

; Sequence 3, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1850 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-724-466B-3

alignment_scores:

Quality: 1738.00 Length: 497
Ratio: 4.080 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.008

alignment_block:

US-09-668-482-4 x US-08-724-466B-3 ..
Align seg 1/1 to: US-08-724-466B-3 from: 1 to: 1850

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
||||| :|||||
88 ATGGGCGTGTACACCCCTTATGCTACACCTTCTCTGCAACCATCGTGTACC 137
17 OLeuLeuLeuPheLeuAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
138 CGTTTACTCTTCTTCGCCCGCGGTGAAGTGTGGAGATGTTAATGATCC 187
34 erGlyArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
188 GACGAGTCGATCCGAACCTGCAGAACTCCTCTACCGCCAGGATCATTGGGC 237
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgGlySph 67
|||||
238 TTGCCGTTTCAATGGAGAAACGCTCCAGCTGATCTCCAGAGAGGAAGTT 287
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
288 TCTGCGCATGAAACGGCAGAAATACGGGTGCATCTACAAAGCACCTCT 337
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
338 TCGGAACCCGCACTGTCAAGGTGATGGAGCTGATTAATGTAGGCAGATT 387
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
388 CTGCTGGCGGCAACACAAGCTGTGTCTGTTCAGTGGCCAGCATCAGTGAG 437
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
438 AACCATCTGGGCTCTGCACACCCCTCTCCAATGTCCATGTGAGTTCAACA 487
134 ySglnArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
488 AAACAAGAAAGAGCCATTATGAGGGCTTCTCTCGAGATGCTCTGGAG 537
151 CysTyValProValIleThrGluGluValGlySerSerLeuGlnIleTr 167
|||||
538 CACTACATTCCTCGATCCAGCAGAGAGTGAAGAGCGCATACAGGAATG 587
167 pleu.....SerCysGlyGluArgGlyLeuLeuValTyrProGluV 181
|||||
588 GCTGCAAAAAGACTCCTGC.....GTGCTGGTTATCCAGAAA 625
181 alLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
|||||
626 TGAAGAAACTCATGTTTCGGATAGCTATGAGAAATCCTGCTGTTTGAA 675
198 Pro...GlnLeuAlaGlyAspGlyAspSerGlnGlnLeuValGluAl 213
|||||
676 CCAGAGCAAAATAAGACGGAC.....GAGCAAGAATCGGTGGAGAC 716
213 aPheGlnGluMetThrArgAsnLeuPheSerLeuProIleAspValProp 230
|||||
717 TTTGAGGAATGATCAAAACTTGTCTCTCTGCAATGCAGCTTCTT 766
230 heSerGlyLeuTyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArg 246
|||||
767 TCAGTGTCTGTACAGGGGTTGAGGGCACGCAATTTCATTCACTCCAAA 816
247 IleGlnGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAl 263
|||||
817 ATTGAGGAATAACATCAGCAAGAAATTCAGAT...GACGACAATGAATA 863
263 aGlyGlnGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpG 280
|||||
864 CGAACAGAAATACAAAGACGCCCTTCAGCTGTGATCGAAGACAGCAGAA 913
280 luArgGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGlu 296
|||||
914 GAAGTGACGAACCTTTTGTGTTGACAGGCGATGAAGAAGACAGTACAGAG 963
297 LeuLeuPheGlyGlyHisGluThrThrAlaSerAlaIleThrSerLeuI 313

181 allysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
:::|||||
626 TGAAGAAACTCATGTTTCGGATAGCTATGAGAAATCTGCTGGTTTGAA 675
198 Pro...GlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAl 213
||| |||
676 CCAGAGCAAAATATAAGACGGAC.....GAGCAAGAACTGCTGGAAGC 716
213 apheGluGluMetThrArgAsnLeuPheSerLeuProIleAspValProp 230
|||||
717 TTTGAGGAAATGATCAAAACTGTTCTCTCTGCAATCGACGTTCTCT 766
230 heserGlyLeuTyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArg 246
|||||
767 TCAGTGGTCTGTACAGGGGTTTGAGGGCAGCAATTTCATTCTCCAAA 816
247 IleGluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAl 263
|||||
817 ATTGAGGAAATATCAGGAAGAAATTCAGAT...GACGACAAATGAAAA 863
263 aglyGlnGlyCysLysAspAlaLeuGlnLeuIleGluHisSerTrpG 280
||| |||
864 CGAACAGAAATACAAAGACGCCCTTCAGCTGTTGATCGAGAACAGCAGAA 913
280 IuarGlyGlyIuarGleuAspMetGlnAlaLeuLysGlnSerSerThrGlu 296
::: |||
914 GAAGTGACGAACCTTTTATTGTTGACAGGCGATGAAAGAAAGCAGCTACAGAG 963
297 LeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuI 313
|||||
964 CTTCTATTGGAGGTCATGAACACCGCCAGCAGCTCAACCTCACTTGT 1013
313 eThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluGlu 330
: |||
1014 CATGTTTCTGGGTGTAACACAGAAAGTGTGACAGAAAGTACAGAGAGAGG 1063
330 eulysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAsp 346
::: |||
1064 TTCAGGAGGAAGGTTGAATGGGCATGTATACACCTGGAAGGGCTTGAAGT 1113
347 MetGluIleLeuGlnLeuLysTyrIleGlyCysValIleLysGluTh 363
|||||
1114 ATGAGAGCTGTGGACAGCTGAAGTACATGATGTGTATTAAGAGAC 1163
363 rLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysT 380
|||||
1164 TCTTAGAATCAACCTCTCTGTTCCCGAGAGATTCAAGTCCGACTCAAAA 1213
380 hrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrPasnValIleTyr 396
|||||
1214 CTTTGAATTTGAATGTTTACCAAAATTCCTAAAGGATGGAACGTCATTAC 1263
397 SerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysGluI 413
|||||
1264 AGCATCTGTGACACGACGATGTGGCCGACGCTCTTCCAAACCAAGAGGA 1313
413 upheAsnProAspArgPheSerAlaProHisProGluAspAlaSerArgP 430
|||||
1314 GTTCCAGCCGAGAGATTATGAGCAAAAGTCTGAGGAGCGGCTCAGGT 1363
430 heserPheIleProPheGlyGlyLysArgSerCysValGlyLysGlu 446
|||||
1364 TTTACTACATCCCTTCGAGAGGATTCAGGATGTGTGGGCAAGAG 1413
447 pheAlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCy 463
|||||
1414 TTCCGCAAAAGTGTACTCAGATCTTTTAGTTGAGTTAAGCAGCAGCATG 1463
463 sAspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrV 480
||: |||
1464 CAATGTGATTTCTCTCAACGAGACCCCGACAAATGAAAAACAGCCCGACTA 1513

480 alTyrProValAspAsnLeuProAlaArgPheThrHisPhe 493
:::|||||
1514 TTTACCAGTGGACAAATCTCCCTACCAAAATTCACATGATTAT 1554

seq_name: /cgn2_6/ptodata/2/lna/6b_COMB.seq:us-08-882-164D-38

seq_documentation block:

; Sequence 38, Application US/08882164D
; Patent No. 6306624

; GENERAL INFORMATION:

; APPLICANT: Petkovich, P. Martin, White, Jay A.,

; APPLICANT: Beckett, Barbara R., Jones, Glenville

; TITLE OF INVENTION: Retinoid Metabolizing Protein

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Blake, Cassels & Graydon

; STREET: Box 25, Commerce Court West

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5L 1A9

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage

; COMPUTER: COMPAQ, IBM PC compatible

; OPERATING SYSTEM: MS-DOS 5.1

; SOFTWARE: WORD PERFECT

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/882, 164D

; FILING DATE: June 25, 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/667, 546

; FILING DATE: June 21, 1996

; APPLICATION NUMBER: 08/724, 466

; FILING DATE: October 1, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Hunt, John C.

; REGISTRATION NUMBER: 36,424

; REFERENCE/DOCKET NUMBER: 50767/00010

; TELEPHONE: (416) 863-4344

; TELEFAX: (416) 863-2653

; INFORMATION FOR SEQ ID NO: 38:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4164 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-882-164D-38

alignment_scores:

Quality: 1712.00 Length: 1046

Ratio: 4.019 Gaps: 10

Percent Similarity: 40.727 Percent Identity: 38.815

alignment_block:
US-09-668-482-4 x US-08-882-164D-38 ..

Align seg 1/1 to: US-08-882-164D-38 from: 1 to: 4164

40 CysAlaLeuProLeuProProGlyThrMetGlyPheProPhePheGlyG 56
|||: |||

73 TGGGTCCTCTCTCCACCTCCCTCT..... 99
|||: |||

56 uThrLeuGlnMetValLeuGlnArgArgLysPheLeuGlnMetLysArg 73
||| |||

100CAGCGAGGAAGTTTCTGCAGATGAAGCGCA 130
|||: |||

73 rglYstYrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVal 89
||| |||

131 GGAATACGGCTTCATCTACAGACGCACTCTTTGGGCGGCCACGCTG 180
||| |||

90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAsp 106
||| |||

|||||
181 CGGTGATGGCGCGGATATGCGCGCATCTTGTGGAGAGCACC 230
106 gLeuValSerValHisTrpProAlaSerValArgThrIleuGlySer 123
|||||
231 GTTGTGTGCTGACATGCGCGCGCTGCGCACCATTCTGGCGCTG 280
123 lYcysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys..Lys.. 138
|||||
281 GCTGCTCTCCAACTGCACGATCTCTGCACAAAGCAGCAAGAAAGGTG 330
138 138
331 AGGTGAGCTGGCAACTCTTGTGCTGGCAGGAGACCTCATCTATGGCT 380
138 138
381 TGGTTCAGGCAAAATAGATGCGGGCAGGGCTAGTCTATGTGTGG 430
138 138
431 GACCAGACCTCTCTATCTGATCCACTTATGCTTCTGCTAGCAGC 480
138 138
481 TGGTTCCTGCTGGGGGAGCTGAAATCTTGAAGGTAAGTCTGGAAGG 530
138 138
531 CGAAGGGGGGGGCTGAGGGAAGTAGAGATTGTAACTCTCTGCTC 580
139 ValIleMetArgAlaPheSerArgGlnAlaLeuGln 150
581 CTGGGGGGTCTCAGGTGATTATGACGGCTTTCAGCGCGGCACTCCA 630
150 uCysTyrValProValIleThrGluGluValGlySerSerLeuGlnI 167
:|||||
631 GTGCTACGTGCGCGTGTATGCTGAGGAAGTCAAGCAGTTGTCTGAGCAGT 680
167 rPLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183
|||||
681 GGCTAAGCTGCGCGGAGCGCGGCTCTGCTACCCCGAGGTGAAGCGC 730
184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnI 200
|||||
731 CTCATGTTCCGATCGCCATGCGCATCTGCTGGCTGCGAGCCGGGTCC 780
200 uAlaGlyAspGlyAspSerGluGlnGlnLeuValGlnAlaPheGluGln 217
|||||
781 AGCGGGCGGGGAGGAGCAGCAGCAGCAGCTCGTGAGGCTTTTCGAGGAGA 830
217 eThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
|||||
831 TGACCCGCAATCTCTCTCTCTCCATTGACGTGCCCTTTAGCGGCTG 880
234 TyrArg:..... 235
881 TACCGGGTAGGGCGGTTTGGCGAGTCGAGTAGGGGAGCAAGCTCGG 930
236 Gly..MetLysAlaArg 240
931 GCATCCGCTCACCGCCACAGCTCTCTCCGCGCTCAGGGCGGTGAAGGCGCG 980
241 AsnLeuIleHisAlaArgIleGluGlnAsnIleArgAlaLysIleCysGln 257
|||||
981 AACCTTATACAGCGCGCATCGAGGAGAACATTCGCGCCCAAGATCCGCGCG 1030
257 yLeuArgAlaSerGlnAlaGlyGlnGlyCysLysAspAlaLeuGlnLeu 274
|||||
1031 GCTTCAGGCTACAGAGCCGAGTGGGGTTGCAAGAGACGCGCTGCAGCTCC 1080
274 euIleGluHisSerTrpGluArgGlyGluArgLeuAspMetGln..... 288
|||||

1081 TGATTGAGCACTCGTGGGAGAGGGAGAGAGCTGTGATATGCAGGTGAGA 1130
288 288
1131 AGCAATTCAAAAGGTGCCAAGGCGCGGAGTGCCTCTGACTTTCAGA 1180
288 288
1181 CACACTTCTGGGCTGCCAAAGCCCTGTCAAGGCCCCAGCTACTTCCA 1230
288 288
1231 GTGGGCGCGGATGCTAGGTCTAGAGCTTTCAACCTGTGGTCTGACCC 1280
288 288
1281 CTTCAAGGAGCCAAACAACCTTTCAGAAAGGTCGCCCTAAGAGCATCTGC 1330
288 288
1331 ATATCCGATATTACATCAAGAAACATTAACAGTAGCAAAATTACCGTTAT 1380
288 288
1381 GAAGTAGCAACAAGATATTATTATCGTTGGGGGTCACCACACACAGAG 1430
288 288
1431 AACCGTATTAAAGGTGCGATGTGCTAGAGAGCTGTGAAGGGGTGCGC 1480
288 288
1481 TGAGCAATGGGGAAGATCCCAAAGTTCAAAGGCAAGCTCATCTACAAA 1530
288 288
1531 GGTAAAGCGGAAGAGCAGATTAAAGGATTTGCGTTTGTGTGG 1580
288 288
1581 TCTTGACTTCTATGAACAACAAACGATTTTACCCTTGAAAGTCTCCGTG 1630
288 288
1631 CAATATTCTCAGGTCAAGTCTTTGTAACAGTCTATAACTGCACACAGA 1680
289 AlaLeuLysGlnSerSerThr 295
1681 TCTGTATAAAGTTCGCTTTTATCCCTTAGGCACTAAACAAATGCTCAACA 1730
296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerI 312
|||||
1731 GAGCTCCTCTTGTGTCATGAAGAACTACAGCAGTGTGCGAGCTCACT 1780
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGlu 329
|||||
1781 GATCACTTACCTAGGACTTACCCACATGCTCCAGAAAGTTGAGAAAG 1830
329 lLeuLysSer..... 332
1831 AGATTAAGAGCAAGGTAGGATGATTCTAGAGTTCCCATTTGCCTAGGA 1880
332 332
1881 CATTCCTCTATTAAACACACACACACCCCACTGTATATAAGTTGCTC 1930
332 332
1931 GATACACCCAGTACTATGACAGTGAAGATCTGAGAGCTAGGTGGAGTGT 1980
332 332
1981 GGGGAGAGACTCCACCTGCTGAATTTAAAGGACAGTTGTTGTAAGTGG 2030

```
332 ..... 332
2031 GCTCTCTTGGGAGAAATTGACCCCTCTCCTCCTCCTCCTCCT 2080
332 ..... 332
2081 CCTCTTCTCTCCACCACCACCACCATCACCACCTTTATAGCAAGG 2130
332 ..... 332
2131 TTCTCTTTCCTGACCAAGAACATGATATGATTAGAGCCAATAGC 2180
332 ..... 332
2181 TGATCAGGGTCGAGTGTGTGAGGGCTCAGGGTATGACCCCTTATATA 2230
332 ..... 332
2231 CCTGATAAGCAACATGTCTGATATAGGCTTAGGCTGAGGAAGTGTGG 2280
332 ..... 332
2281 AAAGGAAGGCCATCAGGCCATCAGCTCTTCCCTTTATCCTCCTCCATC 2330
332 ..... 332
2331 CAGACGCCCTTCAAGTTTACTTAACAGGCTGAGTCTGCTGGGCTGACTTT 2380
333 ..... Lys. 333
2381 TTTTGGAGTGCCAGGGATCCATCAGCTTTTATCTGTTCCATA 2430
334 .GlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluLeu 350
2431 GGGCTTACTTTGCAAGACAAATCAAGACAAAGTTAGACATGGAACCTT 2480
350 euGluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeu 366
2481 TGGCACAGCTTAATATACACTGGGTGTGTCAATTAAGAGACCTGCGAATTG 2530
367 AsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu.. 382
2531 AATCCTCCGGTTCAGAGAGGGTTTCGGGTGCTCTGAAGACTTTTGAGCT 2580
382 ..... 382
2581 GAATGTGAGTGACCTCTCTGTCCCCACCCCGAGCCCTGCTCCAGCTCCA 2630
382 ..... 382
2631 CTCTGCTATGCTGTGACATCAGCTGCCCAGAGCAGTGGCTCAGTGGCC 2680
382 ..... 382
2681 TTGACAGTGTCTGCTGCTCTATGCTACTGGAACCAATTGCTCTCCTCT 2730
382 ..... 382
2731 CTTAATGCCATCCATGCTAGTAATGACTTTTGTGTGCAAGCTCAGGG 2780
383 ..... LeuAsn..GlyTyrG 387
2781 CCGGATGTCAATTCTTAGAATTTTTTTTTTTTAAACAGGGATACC 2830
387 InIleProLysGlyTyrAsnValIleTyrSerIleCysAspThrHisasp 403
2831 AGATCCCAAGGGCTGGAATGTTATTACAGTATCTGTGACACCCACGAT 2880
404 ValAlaGluIlePheThrAsnLysGluGluPheAsnProAspArgPheSe 420
2881 GTGGCAGATATCTTCACTAACAGAGGAATTTAATCCGACCGCTTTAT 2930
```

```
420 rAlaProHisProGlnAspAlaSerArgPheSerPheIleProPheGlyG 437
:::|||||
2931 AGTCCCTCATCCAGAGGATGCTCCCGGTTCAGCTCATTCATTTGGAG 2980
437 IyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLeuLys 453
|||||
2981 GAGGCTTTCGAGCTGTGTAGGCAAGAGTTTGCAAAATTTCTTTAAG 3030
454 IlePheThrValGluLeuAlaArgHisCysAspTyrGlnLeuLeuAsnG 470
|||||
3031 ATATTACAGTGAGCTGGCTAGGCACCTGTGATTGGCAGCTTCTAAATGG 3080
470 yProProThrMetLysThrSerProThrValTyrProValAspAsnLeuP 487
|||||
3081 ACCCTCTACAATGAAGACAGCCCACTGTGTACCTGTGTGACCAATCTCC 3130
487 roAlaArgPheThrHisPheHisGlyGluLeu 497
3131 CTGCAAGATTTTACCACCTTCAGGAGATATC 3162
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-36

seq_documentation_block:

; Sequence 36, Application US/08882164D
; Patent No. 6306624

; GENERAL INFORMATION:

; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5L 1A9

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage

; COMPUTER: COMPAQ, IBM PC compatible

; OPERATING SYSTEM: MS-DOS 5.1

; SOFTWARE: WORD PERFECT

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/882,164D

; FILING DATE: June 25, 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/667,546

; FILING DATE: June 21, 1996

; APPLICATION NUMBER: 08/724,466

; FILING DATE: October 1, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Hunt, John C.

; REGISTRATION NUMBER: 36,424

; REFERENCE/DOCKET NUMBER: 50767/00010

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 863-4344

; TELEFAX: (416) 863-2653

; INFORMATION FOR SEQ ID NO: 36:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2677 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-882-164D-36

alignment_scores:

Quality: 1365.00

Ratio: 4.320

Percent Similarity: 47.235

Length: 669

Gaps: 7

Percent Identity: 46.936

alignment_block:

US-09-668-482-4 x US-08-882-164D-36 ..

Align seg 1/1 to: US-08-882-164D-36 from: 1 to: 2677

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
676 ATGGGGCTCCCGCGCGCTGCTGGCCAGTGGCTCTGCACCTTCGTGCTGCC 725

17 OLeuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTrCysValS 34
|||||
726 GCTGCTGCTCTTCTCTGCTGGCTGCATCAAGCTCTGGACCTGTACTGCTGA 775

34 erGlyArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
776 GCGGCCGACCCGACAGTTGTGCTCCCATTTGCCCGCGGACTATSGGS 825

51 PheProPhePheGlyGluThrLeuGlnMetValLeu.Gln..... 63
|||||
826 TTCCCTTCTTTGGGAAACCTTGACAGATGNTACTNCAGGTAAGGAGGG 875

63 63

876 TGGGGGGGACAGGCTGCTTCCCGGAGCCCGCGGCTGTGGCTTCT 925

63 63

926 GCTGAAGTCCGGGGTAGGCGCCCGGAGGCATGCTATTGCGGCTAGGAG 975

63 63

976 CAGGCTGGCGGGAGCGCGGCTCCCGGMYMCSCTCAMGCSRCRWK 1025

64ArgArgLysPheLeuGlnMetLysArg 72
||| ||| |||
1026 TMMCTCCGCCCTTCTCCCAMAGCGARSARWKYKGMRGATGAAGCGC 1075

73 ArgLysTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVa 89
|||||
1076 AGGAAATACGGCTTACATCTACAGACGACATCTGTTCGGCGGCCACCGT 1125

89 IArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAspA 106
|||||
1126 ACGGCTGATGGCGCGGACAACTGCGGCGCATCTTGCTCGGAGAGACAC 1175

106 rGLeuValSerValHisTrpAlaSerValArgThrIleLeuGlySer 122
|||||
1176 GCGTGGTGTGCGTCCACTGGCCAGCGTCCGTCGCACCATTTCTGGGATCT 1225

123 GlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLys.. 138
|||||
1226 GCGTGCCTCTCTAACCTGCACACTCTCTCGACACAAGCAGCAAGAAAGT 1275

138 138

1276 GGGGGCAGAGGCGAGCGCTGAGACAGGGAGGGGACCCCAATTATGAGCG 1325

138 138

1326 GAATTCCGGCTGATGATGCTAGCGCGGCTAGCAGCTTGAGGTGGGCT 1375

138 138

1376 AGGACCCCTCTGCCAGTCCAGGTTAGCTTTCCAGCTCGAGAGAGTGCCAT 1425

138 138

1426 GTGTCTGGCAGACTGGGGGTGTCTGGAAGGGAGCGGGGTAGACGAGAG 1475

139ValIle 140
|||||
1476 GGGCGGATGAGGCTTTAAAGCTGTCCCTCTCGGGACTCAGGTGATT 1525

141 MetArgAlaPheSerArgGluAlaLeuGluCysTyrValProValIleTh 157
|||||

1526 ATGCGGGCTTTCAGCCCGCAGGACACTCGAATGCTACGTGCCGGTATCAC 1575

157 rGluGluValGlySerSerLeuGluGlnTrpLeuSerCysGlyGluArg 174
|||||
1576 CGAGGAAGTGGGACAGCAGCCTGGAGCAGTGGCTGAGCTGCGGCGAGCGCG 1625

174 lYLeuLeuValTyrProGluValLysArgLeuMetPheArgIleAlaMet 190
|||||
1626 GCCTCCTGGTCTACCCCGAGGTGAAGCGCCTCATGTTCGAATCGCCATG 1675

191 ArgIleLeuLeuGlyCysGluProGlnLeuAlaGlyAspGlyAspSerG 207
|||||
1676 CGCATCTACTGGGCTGCCAAACCCCACTGGCGGGCGAGCGGACTCCGA 1725

207 uGlnGlnLeuValGluAlaPheGluGluMetThrArgAsnLeuPheSerL 224
|||||
1726 GCAGCAGCTGTGTGAGGCGCTTCGAGGAATGACCCGCAATCTCTTCGCG 1775

224 euProIleAspValProPheSerGlyLeuTyrArg..... 235
|||||
1776 TGCCCATCGACGTGCCCTTCAGCGGGCTGTACCGGGTAAAGGCGGCAAC 1825

235 235

1826 GGGCTGCGGACTAGGGGCGCGGACCTGGGCGTCTGCTCACCGCCCGCG 1875

236Gly.MetLysAlaArgAsnLeuIleHisAlaArgIle 247
||| ||| |||
1876 CTCTCTGCGCTCAGGGCATGAAGGCGCGGAACCTCATTTACGCGCGCATC 1925

248 GluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGl 264
|||||
1926 GAGCAGAACATTCGCGCCAAGATCTGGGGCTGCGGGCATCCGAGCGGG 1975

264 yGlnGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluA 281
|||||
1976 CCAGGCTGCAAAAGACGCGCTGACGCTGTGATCGAGCAGCTCGTGGAGA 2025

281 rGlyGluArgLeuAspMetGln..... 288
|||||
2026 GGGGAGACGGCTGACATGACAGTGACTAGACGTTCAAGACAGGCACT 2075

288 288

2076 GCGGAGTTTGGTCCCGCTGGCTTTCCAAAGCGCTGTTCTCGGGCCCCAA 2125

288 288

2126 AGCGCGCGCCTGGGGCCAGCTTCTGAGAGTGGCGCGCGGCTCAGACTA 2175

288 288

2176 CAGCTATGGAATCCCGAAGAAGGCTGAGACACCCGGTCAGAGAGCTGC 2225

288 288

2226 GGAAGGGGCTGCGGMGAACCTGGAGACATCCCCCTAGCCTTAMCAGGTT 2275

288 288

2276 TCAAAAGGAAAGTTGGAATTTCGAAAAATGTTAATAAAGAACTTGCGAT 2325

288 288

2326 TTTAATAAACTAAAGACTTTAACTCAGAGATTTCGGRAGCGGGGTCG 2375

288 288

2376 TACTCGCCTTACTGCTCCAGCTGAACCTAAAGGAGAGCTTGCAATTGTTTA 2425

288 288

2426 AAGATATTGCTTCTCTGACTTCTGTCAAGCAAAACATTTAGCCCTTCTA 2475


```
288 ..... 288
2476 GTCTTCCCTCCAGAACTCTCAGTTGATTTCTGAGTAATCCTTCTGTCAAA 2525
289 .....Ala..Leu 290
2526 CCGCAGGCAGACTGTGTGAAATGTGGGTCTCAGCTTAATCTTAGGCACATA 2575
291 LysGlnSerSerThrGluLeuLeuPheGlyGlyHisGluThrThrAlaase 307
2576 AAGCAATCTTCAACCGAATCTCTTTGGAGACACGAAACACGCGCCAG 2625
307 rAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuG 324
2626 TGCAGCCACATCTCTGATCACTTACCTGGGGCTCTACCCACATGTCTCC 2675
324 ln 324
2676 AG 2677
```

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-724-466B-11

seq_documentation_block:

```
; Sequence 11, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch; 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-724-466B-11
```

```
alignment_scores:
    Quality: 614.00      Length: 117
    Ratio: 5.248        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
```

alignment_block:

US-09-668-482-4 x US-08-724-466B-11 ..

Align seg 1/1 to: US-08-724-466B-11 from: 1 to: 351

```
296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerle 312
1 GAATCTCTCTTTGGAGGACACGAAACCGGCCAGTGCAGCCACATCTCT 50
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluG 329
51 GATCACTTACTTGGGGCTCTACCCACATGTTCTCCAGAAAGTGGCAGAAG 100
329 IuLeuLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 345
101 AGCTGAAGAGTAAGGGTTTACTTTGCAAGACCAATCAAGACACAGATTG 150
346 AspMetGluIleLeuGluGlnLeuLysTyrIleGlyCysValIleLysG 362
151 GACATGGAATTTTGGAAACAACCTTAATACATCGGGTGTATTAAAGGA 200
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuL 379
201 GACCCCTTCAGCTGAAATCCCCAGTTCACAGAGGGTTTCGGGTGCTCTGA 250
379 ySThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIle 395
251 AGACTTTTGAATTAATGATACAGATTCCCAAGGGCTGGAATGTTATC 300
396 TyrSerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysG 412
301 TACAGTATCTGTGATACTCATGATGTGGCAGAGATCTTCACCAACAAGGA 350
412 u 412
351 A 351
```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-882-164D-11

seq_documentation_block:

```
; Sequence 11, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-11

alignment_scores:
Quality: 614.00 Length: 117
Ratio: 5.248 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-668-482-4 x US-08-882-164D-11 ..

Align seg 1/1 to: US-08-882-164D-11 from: 1 to: 351

296 GluLeuLeuphegIyGlyHisGluThrAlaSerAlaAlaThrSerIe 312
1 GAACTCCCTCTTGGAGGACACGAAACCGCCAGTGCCAGCCACATCTCT 50
312 uIleThrTyrLeuGlyLeuTyrProHisValleuGlnLysValArgGluG 329
1 GATCACACTACCTGGGGCTCTACCCACATGTTCTCCAGAAGTCCGAGAAG 100
51 GATCACACTACCTGGGGCTCTACCCACATGTTCTCCAGAAGTCCGAGAAG 100
329 IuLeuLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 345
1 AGCTGAAGAGTAAGCGCTTACTTTCACAGACCAATCAAGACACAACTTG 150
101 AGCTGAAGAGTAAGCGCTTACTTTCACAGACCAATCAAGACACAACTTG 150
346 AspMetGluLeuLeuGluGlnLeuLysTyrIleGlyCysValIleLysG 362
1 GACATGGAAATTTTGAACACACTTAATATCATCGGGTGTATTATTAAGA 200
151 GACATGGAAATTTTGAACACACTTAATATCATCGGGTGTATTATTAAGA 200
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuL 379
1 GACCTTCGACTGAATCCCGCAGTTCAGAGAGGCTTTCGGGTTCCTCGA 250
201 GACCTTCGACTGAATCCCGCAGTTCAGAGAGGCTTTCGGGTTCCTCGA 250
379 yStrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrPasnValIle 395
1 AGACTTTTGAATTAATGATACAGATCCCAAGGCGCTGAATGTATATC 300
251 AGACTTTTGAATTAATGATACAGATCCCAAGGCGCTGAATGTATATC 300
396 TyrSerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysG 412
1 TACAGTATCTGTGATACTCATGATGTGGCAGAGATCTTCACCAACAAGA 350
301 TACAGTATCTGTGATACTCATGATGTGGCAGAGATCTTCACCAACAAGA 350
412 u 412
351 A 351

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-622-166A-1

seq_documentation_block:

Sequence 1, Application US/08622166A
Patent No. 5952545
GENERAL INFORMATION:
APPLICANT: KONCZ, CSABA
APPLICANT: MATHUR, JAIDEEP
APPLICANT: SZEKERES, MIKLOS
APPLICANT: ALTMANN, THOMAS
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,166A
FILING DATE: 27-MAR-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0147-0153P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1608 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: A. thaliana
IMMEDIATE SOURCE:
LIBRARY: lambda gt10
CLONE: C204
FEATURE:
NAME/KEY: CDS
LOCATION: 48..1466
US-08-622-166A-1

alignment_scores:
Quality: 433.50 Length: 486
Ratio: 1.521 Gaps: 13
Percent Similarity: 58.642 Percent Identity: 26.543

alignment_block:
US-09-668-482-4 x US-08-622-166A-1 ..

Align seg 1/1 to: US-08-622-166A-1 from: 1 to: 1608

19 LeuLeupheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValSerG 35
1 GCTCCTCCCTCTCTCCATCGCCCGCGCTTCTCCTCTACTCCGC.. 113
66 GCTCCTCCCTCTCTCTCCATCGCCCGCGCTTCTCCTCTACTCCGC.. 113
35 YArgAspArgSerCysAlaLeuProLeuProGlyThrMetGlyPheP 52
1 GGTACACGTTACCGCTCGGATGGGTCTCGCTCCGGGAAGCCTTGCTCC 162
114 .GGTACACGTTACCGCTCGGATGGGTCTCGCTCCGGGAAGCCTTGCTCC 162
52 rophepheGlyGluThrLeuGlnMetVal.....LeuGln 63
1 GCTGATAGAGAGACTTTTCAGCTGATCGAGCTTACAAACAGAGAAC 212
163 GCTGATAGAGAGACTTTTCAGCTGATCGAGCTTACAAACAGAGAAC 212
64 ArgArgLysPheLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLy 80
1 CCTGAGCCTTTCATCGACGAGAGAGTAGCCCGGTACGGTTCGGTTTCAT 262
213 CCTGAGCCTTTCATCGACGAGAGAGTAGCCCGGTACGGTTCGGTTTCAT 262
80 sThrHisLeupheGlyArgProThrValArgValMetGlyAlaAspAsn 97
1 GACGCATCTTTTGGTGAACCGACGATTTTCTCAGCTGACCCGGAACGA 312
263 GACGCATCTTTTGGTGAACCGACGATTTTCTCAGCTGACCCGGAACGA 312
97 aLArgArgIleLeuLeuGlyAspAspArgLeuValSerValHisTrpPro 113
1 ACCGGTTTGTCTTCAGAACGAAAGGAGCTTTTGTAGTGTCTTATTCCT 362
313 ACCGGTTTGTCTTCAGAACGAAAGGAGCTTTTGTAGTGTCTTATTCCT 362
114 AlaSerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAs 130
1 GCTTCCATTTGTAAACCTTTTGGGGAACACCTCTGCTTCTTATGAAGG 412
363 GCTTCCATTTGTAAACCTTTTGGGGAACACCTCTGCTTCTTATGAAGG 412
130 pSerSerHisLysGlnArgLysLysValIleMetArgAlaPheSerArgG 147
1 TTCTTTGCAATAACGATATGACACTCTCTCACCATGAGCCTTTGCTAATCTT 462
413 TTCTTTGCAATAACGATATGACACTCTCTCACCATGAGCCTTTGCTAATCTT 462

147 luAlaleuGlucSTyrValProVal...lleThrgluGluValGlySer 162
::: ::::: ::::: ::::: ||| ::::: |||
463 CAATCATTAAGACCATCTCATGCTGATATATGACCGGTAGTCCGGTTT 512
163 SerleuGluGlnTrpLeuSerCysGlyluArgGlyLeuLeuValTyrPr 179
::: ||| ::::: ||| ::::: |||
513 AATCTGATCTTGCTCTCT...CGTGTCTCTCATGCA 550
179 ogluValLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyC 196
||| ::::: ||| ::::: ||| ::::: |||
551 AGAAGCCAAAAGATAAGTTGAGCTAACGGTGAAGCAGTTGATGAGCT 600
196 ysgluProGluLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGlu 212
::: ||| ::::: ||| ::::: |||
601 TTGATCCA.....GGGAATGGAGTGAGAGTTTAAGCAA 635
213 AlaPheGluGluMetThrArgAsnLeuPheSerLeuProIleAspValPr 229
::: ::::: ||| ::::: |||
636 GAGTATCTTCTTGTCATCGAAGGCTTCTCTCTCTCTCTCTCTCTCTC.. 683
229 oPheSerGlyLeuTyrArg...GlyMetLysAlaArgAsnLeuIleHisa 245
||| ||| ::::: |||
684 .TTCTCCACCACCTAACCCAAAGCCATCCAAAGCCGAGG..... 722
245 laArgIleGluGlnAsnIleArgAlaLysIleCysGlyLeuArg..... 259
::: ::::: |||
723 ..AAGGTGGCGAGCGCTTGACGGTGTGATGAAGAAAGAGGAGAGAG 770
260 AlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeuLeuIle.. 275
::: ||| ::::: |||
771 GAGGAAGAAGAGCGGAGAGAAAGAAAGATATGCTTGCGGCTTGCTGC 820
276 ...GluHisSerTrpGluArgGlyGluArgLeuAspMetGlnAlaLeuL 291
::: ::::: |||
821 GCGCGATGATGGATTTCGATGAAGAGATTGTGACTTC..... 860
291 ysglnSerSerThrGluLeuLeuPheGlyGlyHisGluThrThraLaser 307
861TTGGTGGCTTTACTTGTGCGCGGTATGAACAACCTCCACG 902
308 AlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuG1 324
||| ::::: |||
903 ATCATGACTCTCGCCGTCAAATTCTCACCGAGACTCTTTAGCTCTTGC 952
324 nLysValArgGluGluLeuLysSerLysGlyLeuLeuCysLysSerAsnG 341
::: ||| ::::: |||
953 TCAACTCAAGGAAGAG.....CATGAAGAAGATTAGGGCAATGAAGAGTG 996
341 lnaSpasnLysLeuAspMetGluIleLeuGluGlnLeuLysTyrIleGly 357
::: ||| ::::: |||
997 ATTCTATATAGTCTTGAATGGAGTGATTACAAGTCAATGCCATTACACAA 1046
358 CysValIleLysGluThrLeuArgLeuAsnProProValProGlyGlyPh 374
||| ::::: |||
1047 TGTGTGGTTAATGAGACGCTAGAGTGCGTTAACATCATCGCGGTGTTT 1096
374 earGValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysG 391
||| ||| ::::: |||
1097 CAGACGTGCAATGACGAGATTGAGATCAAAAGTTATAAATTCCAAAAG 1146
391 lYTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaGluIle 407
||| ::::: |||
1147 GGTGAAGATATCTCATCGTTTAGACGGTTCAATTAGACCCAAACACAC 1196
408 PheThrAsnLysGluGluPheAsnProAspArgPheSerAlaProHisPr 424
||| ::::: |||
1197 TTCAAGATGCTCGCACTTTCACCCCTTGAGATGCGACAGCAACTCGGT 1246
424 ogluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArgS 441
::: |||
1247 AACGACAGGCGCTTCTATATGTGTTCACACCGCTTGTGAGGCGCAAGGC 1296

441 erCysValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrVal 457
||| ||| ::::: |||
1297 TATGTCCCGGTACAGAGCTGCTAGGGTTGACACTCTGTGTTTCTTCAC 1346
458 GluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrMe 474
::: ||| ::::: |||
1347 CGCCTAGTACAGAGCTTCAGTTGG.....GTTCTGCAGA 1381
474 tLysThrSerProThrValTyrProValAspAsnLeuProAlaArgPheT 491
::: ::::: |||
1382 GCAAGACAAGCTGTTTCTTTCCTCACTACAGAACGCAAGAAACGTACC 1431
491 hrHisPhe 493
|||
1432 CGATCTTC 1439

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-991-677-3

seq_documentation_block:

; Sequence 3, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent L
; APPLICANT: Carraway, Daniel T
; APPLICANT: Smeltzer, Richard H
; TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
; FILE REFERENCE: 50617
; CURRENT APPLICATION NUMBER: US/08/991,677A
; CURRENT FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: US 60/033,381
; EARLIER FILING DATE: 1996-12-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1883
; TYPE: DNA
; ORGANISM: Liquidambar styraciflua
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (74)..(1606)
US-08-991-677-3

alignment_scores: Quality: 348.50 Length: 507
 Ratio: 1.267 Gaps: 22
Percent Similarity: 54.241 Percent Identity: 26.824

alignment_block:
US-09-668-482-4 x US-08-991-677-3 ..

Align seg 1/1 to: US-08-991-677-3 from: 1 to: 1883

14 PheValLeuProLeuLeuLeuPheLeuAlaIleLysLeuTrpAspLe 30
||| ::::: |||
125 TTCATTATACCTTGTGCTACTCTTATGTGGCCTAGTATCTCGGCTT..... 169
30 uTyrCysValSerGlyArgAspArgSerCysAlaLeuProLeuProProG 47
||| ::::: |||
170CGCCAGAGA.....CTACCATACCCACAG 194
47 lYThrMetGlyPheProPhePheGlyGluThrLeuGlnMetValLeuGln 63
||| ::::: |||
195 GCCCAAAAGGCTTACCGGTGATCGGAACATGCTCATGATGATCACTC 244
64 ArgArgLysPheLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLy 80
::: ::::: |||
245 ACTCACCGAGAGACTCGCAAACTCGCCAAACAATAACGGCGGTCTATTCCA 294
80 sThrHisLeuPheGlyArgProThrValArgValMetGlyAlaAspAsnV 97
::: |||
295 CCTCAAGATGGGATGCTTACACATGTGTGGCGGTTTCCACACCGGACATGG 344

97 aAlaArgIleLeuLeuGlyAspAspArgLeuValSerValHisTrpPro 113
:::||||:::||||:::||||:::||||:::||||:::||||:::||||
345 CTGCCAAGTCTTCAAGTCCAGACAACATCTTCTCG..AACGGCCA 391
114 AlaSerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeu...Hi 129
||||:::||||:::||||:::||||:::||||:::||||:::||||
392 GCCACCATA.....GCCATCAGCTACCTCACCTA 420
129 sAspSerSer.....HisLysG 135
:||||:::||||:::||||:::||||:::||||:::||||:::||||
421 TGACCGAGCCGACATGGCCTTCCTACTACGGCCCGTTTGGCGTCAGA 470
135 InArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlyCys 151
:|||||||:::||||:::|||||||:::|||||||:::|||||||
471 TGGGTAACCTCTGCGTCATGAATTTATTAGCCGGAACGAGCCGAGTCG 520
152 Tyr.....ValProValIleThrG 158
:::||||:::||||:::||||:::||||:::||||:::||||:::||||
521 TGGGAGTCGGTCCGAGACGAGTCGACTCGGCAGTACGAGTGGTCGGTC 570
158 uGluValGlySerSerLeuGluInTrpLeuSerCysGlyGluArgGlyL 175
:|||||||:::|||||||:::|||||||:::|||||||:::|||||||
571 CAATATTGGGTCGACG.....GTGAATATCGGGCAG..... 601
175 euLeuValTyrProGluValLysArgLeuMetPheArgIleAlaMetArg 191
|||||||:::|||||||:::|||||||:::|||||||:::|||||||
602 ..CTGGTTTGTCTCTGACGAGAATATTACTTACAGGGCGGCTTTTGGG 649
192 IleLeuLeuGlyCysGluProGluLeuAlaGlyAspGlyAspSerGluG 208
:::||||:::||||:::||||:::||||:::||||:::||||:::||||
650 ACGATC.....TCGCATGAGGACCAAGA 672
208 nGluLeuValGluAlaPheGluGluMetThrArg.....AsnLeuPheS 223
:|||||||:::|||||||:::|||||||:::|||||||:::|||||||
673 CGAGTTCGTGGCCATACTGCAAGAGTTTTCGCAGCTGTTGGTCTTTTA 722
223 erLeuProIleAspValProPhe.....SerGlyLeu... 233
:||||:::||||:::||||:::||||:::||||:::||||:::||||
723 ATATAGCTGATTTTATCCCTGGCTCAAAATGGGTTCTCAGGGGATTAAC 772
234 TyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAs 250
||| |||||:::||||:::||||:::||||:::||||:::||||
773 GTGAGGCTCAACAAGGACGAGGGCGCTTGATGGTTATTGACAAGAT 822
250 nIleArgAlaLysIle....CysGlyLeuArgAlaSerGluAlaGlyG 265
||| |||:::||||:::||||:::||||:::||||:::||||
823 CATCGACGATCATATACAGAAAGGAGATAAAAACTCGAGAGGTTGATA 872
265 InGlyCysLysAspAlaLeuGluLeuLeuIleGluHisSerTrpGluArg 281
||| |||:::||||:::||||:::||||:::||||:::||||
873 CTGATATGCTAGATGATTACTTCTTTTACGGGTGAGGAGCAAAAGTA 922
282 GlyGluArgLeuAspMetGln..AlaLeuLys..... 291
:|||||||:::|||||||:::|||||||:::|||||||:::|||||||
923 AGCGAATCTGACGATCTTCAAAATTCCATCAAACTCACCAAGACACAT 972
292 .GlnSerSerThrGluLeuLeuPheGlyGlyHisGluThrThrAlaSerA 308
:||||:::||||:::||||:::||||:::||||:::||||:::||||
973 CAAAGCTATCATGACGTAATGTTTGGAGGACCGAAACGGTGGCGTCCG 1022
308 IaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuGln 324
|| |||||:::|||||||:::|||||||:::|||||||:::|||||||
1023 CGATTGAATGGCCATGACGAGCTGATGAAAAAGCCAGAAATCTTAAG 1072
325 LysValArgGluGluLeuLys...SerLysGlyLeuLeuCysLysSerAs 340
|||||||:::|||||||:::|||||||:::|||||||:::|||||||
1073 AAGGTCCAACAAGAACTCGCCGTGGTGGTGGTCTT..... 1108
340 nGlnAspAsnLysLeuAspMetGluIleLeuGluGluInLeuLysTyrIleG 357
||||:::||||:::||||:::||||:::||||:::||||:::||||
1109 ...GACCGGCGAGTCGAAGAGAAAGACTTCGAGAAGCTCACCTACTTGA 1154
357 LysCysValIleLysGluThrLeuArgLeuAsnProProValProGlyGly 373

|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1155 AATGCGTACTGAAGGAAGTCTTCGCTCCACCCACCCATCCCTCCTC 1204
374 PheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLy 390
:||||:::||||:::||||:::||||:::||||:::||||:::||||
1205 CTCCACGAGACTGCCGAGGACGCCGAGGTCGGCGGCTACTACATTCCGGC 1254
390 sGlyTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaGluI 407
:||||:::||||:::||||:::||||:::||||:::||||:::||||
1255 GAAATCGCGGTGATGATCAACGCGTCCGATCGCCGGGAGACAAGAACT 1304
407 lePheThrAsnLysGluGluPheAsnProAspArgPheSerAlaProHis 423
:||||:::||||:::||||:::||||:::||||:::||||:::||||
1305 CGTGGCCGACCCAGATACGTTTAGGCCCTCCAGGTTTCTCAAGACGGGT 1354
424 ProGluAsp.....AlaSerArgPheSerPheIleProPheGlyGlyG 438
||| |||||:::|||||||:::|||||||:::|||||||:::|||||||
1355 GTGCCCGATTTCAAAGGAACAACACTCGAGTTCATCCCATTCGGGTCAGG 1404
438 yLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLeuLysIleP 455
| |||||:::|||||||:::|||||||:::|||||||:::|||||||
1405 TCGTCGGTCTTGCCCCGATATGCAACTCGGACTCTACGCGCTAGACAG. 1453
455 heThrValGluLeuAlaArgHisCys...AspTrpGlnLeuLeuAsnGly 470
||||| |||||:::|||||||:::|||||||:::|||||||
1454 ..ACTGTGGCTCACCTCTTACATGTTTACAGTGGAGAGTTG..... 1492
471 ProProThrMetLysThrSer 477
||| |||||:::|||||||:::|||||||:::|||||||
1493 CCGGACGGGATGAACCGAGT 1513

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103, 840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

alignment_scores:
Quality: 344.00 Length: 446
Ratio: 1.376 Gaps: 12
Percent Similarity: 56.054 Percent Identity: 25.112

alignment_block:
US-09-668-482-4 x US-09-103-840A-2/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765
32 CysValSerGlyArgAspArgSerCysAlaLeuProLeuProProGlyTh 48
||| ||| |||||:::|||||||:::|||||||:::|||||||
4120837TGGGTTCCGCGGAGACACGCGCGATGCTTTGGGGCGGCCACCGGGCT 4120788

147 CACGAGGTGGCCATCATAGCAACATGTTAATGATGACCAACTCACC 196
65 ArgLysPheLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysTh 81
197 CACCGTGGTTAGCCAATTAGCTAAAGATATGGCGGATTTGCCATCT 246
81 HisLeuPheGlyArgProThrValArgValMetGlyAlaAspAsnVala 98
247 CCGCATGGGATTCCTTCATATGTACCGTGTCTCATCACCAGGTGGCTC 296
98 rGArgIleLeuLeuGlyAspAspArgLeuValSerValHisTrpProAla 114
297 GACAAGTCTTCAAGTCCAAGACAGCGTCTCTCG..AACCGGCTTCCA 343
115 SerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeu..HisAs 130
344 ACTATA.....GCTATAAGCTATCTGACTTACGA 372
130 pSerSer.....HisLysGlnA 136
373 CCGAGCGGACATGGCTTTCGCTCACTACGAGCCGTTTGGAGACAGATGA 422
136 rGlyLysValIleMetArgAlaPheSerArgGluAlaLeuGluCysTyr 152
423 GAAAGTGTGTGCATGAAGGTGTTAGCCGTAAAGAGAGCTGAGTCATGG 472
153 ValProVal.....IleThrGluGluValGlySerSe 163
473 GCTTCAGTTCGTGATGAAGTGGAACAATGGTCCGTCGCTCTGTGTAA 522
163 rLeuGluGlnTrpLeuSerCysGlyGluArgGlyLeuLeuValTyrProg 180
523 CGTTGGTAAGCCTATAACGTCGGGAGCAA.....ATTTTGCAC 563
180 luvAllysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCys 196
564 TGACCCGCAACATAACTTACCGGGCAGCGTTTGGG....TCAGCCTGC 607
197 GluProGlnLeuAlaGlyAspGlyAspSerGluGlnLeuValGluAl 213
608 GAG.....AAGGACAAGACGAGTTCATAGAAT 636
213 apheGluGluMetThrArg.....AsnLeuPheSerLeuProIleAspV 228
637 CTTACAAGAGTCTCTAAGCTTTTGGAGCCTTCAACGATGACGGATTTC 686
228 alProPheSerGlyLeuTyr.....ArgGlyMet 237
687 TACCATATTTCCGGTGATCGATCCGCAAGGATAAACAAGCGGCTCGTG 736
238 LysAlaArgAsn.....LeuIleHisAlaArgIleGluGlnAs 250
737 AAGGCCGTAATGATCTAGACGGATTTATGACGATATTTATCGATGAACA 786
250 nileArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGly..... 264
787 TATGAAGAAGAAGAG..AATCAAAAACGCTGTGATGATGGGATGTTG 833
265GlnGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSer 278
834 TCGATACCGATATGTTGATGATCTTCTTGTCTTTTACAGTGAAGAGGCC 883
279 TrpGluArgGlyGluArgLeuAspMetGln..... 288
884 AAATTAGTCAGTGACAGACGCGATCTCAAAATTCATCAAACTTACCCG 933
289AlaLeuLysGlnSerSerThrGlnLeuLeuPheGlyGlyHisGluT 304
934 TGACAATATCAAAAGCAATCATCATGACGCTTATGTTTGGAGGAACGAAA 983
304 hrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrPro 320

984 CGGTAGCGTCGGCGATAGAGTGGGCTTAAACGAGTTATTACGGAGCCCC 1033
321 HisValLeuGlnLysValArgGluGluLeu..LysSerLysGlyLeuLe 336
1034 GAGGATCTAAACGGGTCCACACAGAAGCTCCGGAAGTCTGTGGACTT.. 1081
336 uCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeuGluGln 353
1082GACAGACGAGTTGAAGAATCCGACATCGAGAAGT 1115
353 euLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAsnProPro 369
1116 TGACTTATCTCAATGCACACTCAAGAAACCCTAAGGATGCACCCACCG 1165
370 ValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTy 386
1166 ATCCCTCTCTCTCTCCACGAAACCGCGAGAGACACTAGTATCGACGGTTF 1215
386 rGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAspThrHisA 403
1216 CTTCAATCCCAAGAAATCTCGTGTGATGATCAACAGCGTTTGGCATAGGAC 1265
403 spValAlaGluIlePheThrAsnLysGluGluPheAsnProAspArgPhe 419
1266 GCGACCAACCTCTTGAGCTGACCCGACACGTTTAGACCATCGAGCTTT 1315
420 SerAlaProHisProGluAsp.....AlaSerArgPheSerPheIlePr 434
1316 TTGGAACCGGGCGCTACCGGATTTCAAGGAGCAATTTGAGATTATACC 1365
434 opheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleL 451
1366 GTTCGGGTGCGGTCTAGATCGTGCCCGGGTATGCAA..... 1402
451 euLeuLysIlePheThrValGluLeuAla.....ArgHisCys 463
1403 ..CTAGGTTTATACGCGCTTGACTTAGCCGTGCTCATATATTACATTGC 1450
464 ..AspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSer 477
1451 TTCACGTGGAATTA.....CCTGATGGGATGAAACCAAGT 1486
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2
seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103, 840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
alignment_scores:
Quality: 400.00 Length: 486
Ratio: 1.481 Gaps: 21

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:25:25 ; Search time 89.59 Seconds
(without alignments)
336.311 Million cell updates/sec

Title: US-09-668-482-32

Perfect score: 2612
Sequence: 1 MGIPALLASALCTFVLPDLL.....PTVYPVDNLPARFTYFGDI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_0601:*

1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB	ID	Description
1	2612	100.0	497	19	AAW37735	Cytochrome P450RAI
2	2612	100.0	497	19	AAW44161	Mouse retinoid met
3	2437	93.3	497	19	AAW37734	Human cytochrome P
4	2437	93.3	497	19	AAW44160	Human retinoid met
5	1734	66.4	492	19	AAW37733	Cytochrome zp450RA
6	1734	66.4	492	19	AAW44159	zebrafish retinoid
7	1052	40.3	216	21	AAW12489	Human PSEC64 prote
8	474	18.1	468	21	AAW20783	Arabidopsis thalia
9	461	17.6	513	21	AAW07921	A cytochrome P450
10	455.5	17.4	461	21	AAW6490	Arabidopsis thalia
11	455.5	17.4	462	21	AAW46489	Arabidopsis thalia

12	455.5	17.4	465	21	AAW46491	Arabidopsis thalia
13	450	17.2	472	18	AAW27153	Arabidopsis thalia
14	450	17.2	472	21	AAW44571	Arabidopsis thalia
15	450	17.2	472	21	AAW45022	Arabidopsis thalia
16	450	17.2	491	21	AAW45021	Arabidopsis thalia
17	450	17.2	492	21	AAW44570	Arabidopsis thalia
18	445	17.0	444	21	AAW44572	Arabidopsis thalia
19	445	17.0	444	21	AAW45023	Arabidopsis thalia
20	440.5	16.9	430	21	AAW20784	Arabidopsis thalia
21	423.5	16.2	461	21	AAW11836	Arabidopsis thalia
22	423.5	16.2	462	21	AAW11835	Arabidopsis thalia
23	423.5	16.2	465	21	AAW11834	Arabidopsis thalia
24	406	15.5	481	21	AAW30049	Arabidopsis thalia
25	406	15.5	489	21	AAW30048	Arabidopsis thalia
26	380.5	14.6	433	21	AAW30050	Arabidopsis thalia
27	380	14.5	475	21	AAW07676	Arabidopsis thalia
28	358.5	13.7	511	20	AAW23341	A P450-2 protein i
29	358.5	13.7	511	21	AAW19694	Sweetgum coniferyl
30	350.5	13.4	513	22	AAW31008	Amino acid sequenc
31	350.5	13.4	513	22	AAW48181	B. napus F5H polyp
32	347	13.3	388	21	AAW20785	Arabidopsis thalia
33	342	13.1	520	22	AAW31007	Amino acid sequenc
34	342	13.1	520	22	AAW48179	B. napus F5H polyp
35	341	13.1	520	22	AAW48180	B. napus F5H polyp
36	337	12.9	520	18	AAW26640	Arabidopsis thalia
37	337	12.9	520	19	AAW40099	Arabidopsis ferula
38	337	12.9	520	21	AAW15188	Arabidopsis ferula
39	333.5	12.8	163	21	AAW40557	Human ORFX ORF321
40	314.5	12.0	520	21	AAW23917	Arabidopsis thalia
41	311.5	11.9	526	20	AAW05675	Maize ferulate-5-h
42	311	11.9	471	21	AAW23013	Arabidopsis thalia
43	311	11.9	479	21	AAW23012	Arabidopsis thalia
44	308.5	11.8	469	21	AAW23014	Arabidopsis thalia
45	300	11.5	576	20	AAW09190	Soybean cytochrome

ALIGNMENTS

RESULT 1

AAW37735 1
ID AAW37735 standard; Protein; 497 AA.

XX AAW37735;

XX AC

XX DT 07-JUL-1998 (first entry)

XX DE

XX DE Cytochrome P450RAI isoform.

XX DE

XX KW Retinoid regulated gene; cytochrome P450 gene; enzyme;

XX KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter; isoform.

XX OS

XX OS Synthetic.

XX PN

XX PN WO9749832-A2.

XX PD

XX PD 31-DEC-1997.

XX PF

XX PF 23-JUN-1997; 97WO-CA00488.

XX PR

XX PR 01-OCT-1996; 96US-0724466.

XX PR 21-JUN-1996; 96US-0667546.

XX PA

XX PA (TOOH) UNIV QUEENS KINGSTON.

XX PI

XX PI Petkovich PM;

XX DR

XX DR WPI; 1998-077193/07.

XX DR N-PSDB; AAV09252.

XX PT

XX PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and

PT psoriasis
XX
PS Disclosure; Pages 596-59H; 113pp; English.
XX
CC This amino acid sequence is of an isoform of cytochrome zp450RAI.
CC Its expression is dependent on the presence of retinoic acid (RA).
CC The retinoid-regulated genes such as the inducible cytochrome p450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC p450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome p450
CC nucleotide sequence can be used to induce or suppress the expression
CC of its protein. P450RAI is highly induced by RA in cell lines and
CC tissues. This allows for development of a drug screen using promoters
CC and nucleotide sequences to identify drugs which are useful for
CC reducing the catabolism of RA.

XX Sequence 497 AA;

Query Match 100.0%; Score 2612; DB 19; Length 497;
Best Local Similarity 100.0%; Pred. No. 9e-244;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPALASALCTFVLPDLLFLAALKLMDLYCVSSRDSCALPLPGTMGPFEGETLQ 60
DB 1 mglpallasalctfvlplllflaalklwdlycvssrdscalpplpgtmgpfifgetlqm 60
QY 61 VLQRRKFLQMKRRKRYGFYKTHLFGPPTVRYMGADNVRRIILGHRVSVHWPASVRTIL 120
DB 61 vlqrrkflgmkrkygfikytlhfgprptvrymgadnvrillghrvsvhwpasvrtill 120
QY 121 GAGCLSNLHDSHKKRKVIMQAFSREALQCVLVIAEVSCLQEWLSCGERGLVYPE 180
DB 121 gagclsnlhdsnhkrrkvimqafsréalqcyvlviaevsscleqwlscgergllyype 180
QY 181 VKRLMFRIAMRIILGCEPAGGGEDEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGVKAR 240
DB 181 vkrlmfriamrillgcepaggedeqolveafeemtrnlfsldvpfsglyrgvkar 240
QY 241 NLIHARIEENIRAKTRRLQATEPDGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
DB 241 nlharieeniraktrrlqatepdgckdalqlliehswgerldmqalkgstellfg 300
QY 301 GHETTASAATSLITYLGLYPHVLQKVREETKSKGLLCKSNQDNKLDMETLEQLKYIGCVI 360
DB 301 ghettsaatstlitylglpvhvlokvreetsksgllcksnqdnkldmetleqlkyigcvi 360
QY 361 KETLRNLNPPVPGGFRVALKTFELNGYQIPKGMNVITYSICDTHVADIFTNKEEFNPDREI 420
DB 361 ketlrnlppvpggfrvalktfelngyqipkgnvitysictdhvadiftnkeefnpdrfi 420
QY 421 VPHPEDASRFSPFEGGGLRSCVCKEFAKILLKFTVELARHCDQNLNGPPTMTKSPV 480
DB 421 vphpedasrfspfegglrscvckefakillkftvelarhcdqnlngpptmtksptv 480
QY 481 YPVDNLPAFETTFYQGD 497
DB 481 ypvdnlpafettyfgdi 497

RESULT 2
AAW44161
ID AAW44161 standard; Protein; 497 AA.
XX
AC AAW44161;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolising protein mp450RAI.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antibody; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;

KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening.
XX
OS Mus musculus.
XX
PN W09749815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.

XX Beckett BR, Jones G, Petkovich PM, White JA;
PI
XX WPI; 1998-077178/07.
DR N-PSDB; AAV12205.
XX

PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
PS Claim 1; page 65-66; 110pp; English.

CC This protein comprises a novel mouse retinoid metabolising protein,
CC designated mp450RAI. Its amino acid sequence was deduced from a
CC cDNA clone (see AAV12205) isolated from a retinoic acid-treated P19
CC teratocarcinoma library. It includes a haem-binding motif
CC characteristic of cytochrome P450 proteins. mp450RAI is a retinoid
CC oxidase that has the ability to hydroxylate retinoic acid at the 4
CC position of the beta-ionone ring, and is inducible in epithelial
CC cells exposed to retinoic acid. zebrafish, human and mouse P450RAIs
CC (see AAW44159-61) are claimed. They can be expressed in host cells
CC and used to metabolize retinoic acid in an organism or cell, in drug
CC screening, and to raise antibodies useful for inhibiting retinoic
CC acid hydroxylation for the treatment of cancer, actinic keratosis,
CC oral leukoplakia, secondary tumours of the head and/or neck,
CC non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis.

XX Sequence 497 AA;

Query Match 100.0%; Score 2612; DB 19; Length 497;
Best Local Similarity 100.0%; Pred. No. 9e-244;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPALASALCTFVLPDLLFLAALKLMDLYCVSSRDSCALPLPGTMGPFEGETLQ 60
DB 1 mglpallasalctfvlplllflaalklwdlycvssrdscalpplpgtmgpfifgetlqm 60
QY 61 VLQRRKFLQMKRRKRYGFYKTHLFGPPTVRYMGADNVRRIILGHRVSVHWPASVRTIL 120
DB 61 vlqrrkflgmkrkygfikytlhfgprptvrymgadnvrillghrvsvhwpasvrtill 120
QY 121 GAGCLSNLHDSHKKRKVIMQAFSREALQCVLVIAEVSCLQEWLSCGERGLVYPE 180
DB 121 gagclsnlhdsnhkrrkvimqafsréalqcyvlviaevsscleqwlscgergllyype 180
QY 181 VKRLMFRIAMRIILGCEPAGGGEDEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGVKAR 240
DB 181 vkrlmfriamrillgcepaggedeqolveafeemtrnlfsldvpfsglyrgvkar 240
QY 241 NLIHARIEENIRAKTRRLQATEPDGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
DB 241 nlharieeniraktrrlqatepdgckdalqlliehswgerldmqalkgstellfg 300
QY 301 GHETTASAATSLITYLGLYPHVLQKVREETKSKGLLCKSNQDNKLDMETLEQLKYIGCVI 360

Db 301 ghettaaatlitylglyphvlqkvreeikskgllcksnqdnklmeltleqlykycvi 360

QY 361 KETLRINPVPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVADIFTNKEEFNPDREI 420

Db 361 ketlrlnpvpvggfrvalktfelngyqipkgwnvlysicdthdvadiftnkeefnprfi 420

QY 421 VPHPEDASRFSFIPEGGILRSCVKGKEFAKILKIFTVELARHCDWQLNGPPTMKTSPV 480

Db 421 vphpedasrfsfipggilrscvkgkefakillkiftvelarhcdwqlngpptmktspv 480

QY 481 YPVDNLPARFTYFOGDI 497

Db 481 ypvdnlparrfityfgdi 497

RESULT 3
AAW37734
ID AAW37734 standard; Protein: 497 AA.

XX AC AAW37734;
XX DT 07-JUL-1998 (first entry)

XX DE Human cytochrome P450RAI protein.
XX KW Retinoid regulated gene; cytochrome P450 gene; enzyme;
XX KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter.
XX OS Homo sapiens.
XX PN WO9749832-A2.
XX PD 31-DEC-1997.
XX PF 23-JUN-1997; 97WO-CA00488.
XX PR 01-OCT-1996; 96US-0724466.
XX PR 21-JUN-1996; 96US-0667546.
XX PA (TOOH) UNIV QUEENS KINGSTON.
XX PI Petkovich PM;
XX DR WPI: 1998-077193/07.
XX DR N-PSDB: AAV09247.

PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis

PS Example 4; Pages 58-59; 113pp; English.

XX CC This is the amino acid sequence of the human cytochrome P450RAI. Its
XX CC expression is dependent on the presence of retinoic acid (RA). The
XX CC retinoid-regulated genes such as the inducible cytochrome P450RAI
XX CC gene specifically metabolises a derivative of the RA. The cytochrome
XX CC P450 gene in general produces enzymes involved in the oxidative
XX CC metabolism of endogenous and exogenous compounds. The cytochrome
XX CC P450 nucleotide sequence can be used to induce or suppress the
XX CC expression of its protein. P450RAI is highly induced by RA in cell
XX CC lines and tissues. This allows for development of a drug screen
XX CC using promoters and nucleotide sequences to identify drugs which are
XX CC useful for reducing the catabolism of RA.

SQ Sequence 497 AA;

Query Match 93.3%; Score 2437; DB 19; Length 497;
Best Local Similarity 93.4%; Pred. No. 7.3e-227;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGLPALASALCTEVLPLLLFLAALKLMDLYCVSSRDRSCALPLPGTMGFPFGETLQ 60

Db 1 mglpallasaalcctfvlp1llflaalklwdlycvsgdrscalplpgtmgfpfgetlqm 60

QY 61 VLQRRKFLQMKRRKRYGYFIYKTHLFGRPTRVMGADNVRRIILGHRLVSVHWPASVRIIL 120

Db 61 vlqrrkflqmkrrkrygyfiythlfgprtrvmgadnvrriilgddrlsvshwpasvrtll 120

QY 121 GAGCLSNLHDSHKQKKVIMQAFSREALQCYVLVIAEVSQCLEQWLSCGERGLVYPE 180

Db 121 gsgclsnlhdschkqrkvimrafsrealcyvpviteevgsleqwlscgergllyype 180

QY 181 VKRLMFRIAMRILLGCEPPGAGGEDQQLVEAFEEMTRNLFSLPIDVPFSGLYRGVKAR 240

Db 181 vkrlmfriamrillgceppaggedqqlveafeemtrnlfsipidvpfsglyrgvmkar 240

QY 241 NLIHARIEENIRAKIRRLQATEPPDGCCKDALQLLIEHSWGERLDMQALKOSTELLFG 300

Db 241 nlharieenirakirrlqateppdgckdalqllehswergerldmqalkqstelli fg 300

QY 301 GHETTASAATSLITYLGLYPHVLQKVREIKSKGLLCKSNQDNKLDMETLEQLKYGCVI 360

Db 301 ghettaaatstlitylglyphvlqkvreeikskgllcksnqdnklmeltleqlykycvi 360

QY 361 KETLRINPVPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVADIFTNKEEFNPDREI 420

Db 361 ketlrlnpvpvggfrvalktfelngyqipkgwnvlysicdthdvadiftnkeefnprfi 420

QY 421 VPHPEDASRFSFIPEGGILRSCVKGKEFAKILKIFTVELARHCDWQLNGPPTMKTSPV 480

Db 421 aphpedasrfsfipggilrscvkgkefakillkiftvelarhcdwqlngpptmktspv 480

QY 481 YPVDNLPARFTYFOGDI 497

Db 481 ypvdnlparrfityfgdi 497

RESULT 4
AAW44160
ID AAW44160 standard; Protein: 497 AA.

XX AC AAW44160;
XX DT 22-JUN-1998 (first entry)

XX DE Human retinoid metabolising protein hp450RAI.
XX KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
XX KW retinoic acid; human; inhibitor; antibody; cancer;
XX KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
XX KW non-small cell lung carcinoma; basal cell carcinoma;
XX KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
XX KW ichthyosis; therapy; diagnosis; screening.

XX OS Homo sapiens.
XX PN WO9749815-A1.
XX PD 31-DEC-1997.
XX PF 23-JUN-1997; 97WO-CA00440.
XX PR 01-OCT-1996; 96US-0724466.
XX PR 21-JUN-1996; 96US-0667546.
XX PA (TOOH) UNIV QUEENS KINGSTON.
XX PI Beckett BR, Jones G, Petkovich PM, White JA;
XX DR WPI: 1998-077178/07.
XX DR N-PSDB: AAV12204.
XX PT Retinoid metabolising protein - useful to develop products to treat,

PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 1; Page 54-55; 110pp; English.
XX
CC This protein comprises a novel human retinoid metabolising protein,
CC designated hp450RAI. Its amino acid sequence was deduced from a
CC cDNA clone (see AAV12204) isolated from a retinoic acid-treated NT2
CC cell library. It includes a haem-binding motif characteristic of
CC cytochrome P450 proteins. hp450RAI is a retinoid oxidase that has
CC the ability to hydroxylate retinoic acid at the 4 position of the
CC beta-ionone ring, and is inducible in epithelial cells exposed to
CC retinoic acid. Zebrafish, human and mouse P450RAIs (see AAW44159-61)
CC are claimed. They can be expressed in host cells and used to
CC metabolize retinoic acid in an organism or cell, in drug screening,
CC and to raise antibodies useful for inhibiting retinoic acid
CC hydroxylation for the treatment of cancer, actinic keratosis, oral
CC leukoplakia, secondary tumours of the head and/or neck, non-small
CC cell lung carcinomas, basal cell carcinomas, acute promyelocytic
CC leukaemia, skin cancer, and premalignancy associated with actinic
CC keratosis, acne, psoriasis and/or ichthyosis.
XX
SQ Sequence 497 AA;

Query Match 93.3%; Score 2437; DB 19; Length 497;
Best Local Similarity 93.4%; Pred. No. 7.3e-227;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPILLFLAALKLMDLYCVSSRDSCALPLPPGTMGFPFEGETLQM 60
DB 1 mglpallasalctfvlpillflaalklwdlycvsgdrscalpippgtmgfpfgetlqm 60
QY 61 VLQRRKFLQMKRRKYGFITYKTHLFGPRTVRVMGADNVRRILLGEHRLVSVHWPASVRTL 120
DB 61 vlqrrkflmqmrrkygfiythlfgprtvrvmgadnvrrillgddrlvsvhwpasvrtll 120
QY 121 GAGCLSNLHDSHKKRKVKVIMQAFSREALQCYLVLYAEVSSCLEQWLSGGERGLVYPE 180
DB 121 gsgclsnlhdsdhkrrkvkvimraqfsrealqcyvlvlyaevvsscleqwlsggergllype 180
QY 181 VKRLMFRIAMRILGCEPBPAGGGEDEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGVKAR 240
DB 181 vkrlmfriamrillgcebpagggedeqolveafeemtrnlfslpidvpfsglyrgvmkar 240
QY 241 NLIHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWEMGERLDMQALKOSTELLFG 300
DB 241 nlharieenirakirrlqatepddggckdalqllliehswemgerldmqalkgstellfg 300
QY 301 GHETTASAATSLITYLGLYPHVLOKVVREIRKSKGLCKSNODNKLDMETLEQLKYTCVI 360
DB 301 ghettasaatslitylglyphvlokvvreelkskglcksnodnklhmeileqlkyigcvi 360
QY 361 KETLRINPVPVPGGFRVALKTFELNGYQIPKGMNVYISICTHADVADIFTNKEEFNPDREI 420
DB 361 ketlrlnppvpvggfrvalktfelngyqipkgmnvyisicdthadvadiftnkeefnpdrfi 420
QY 421 VPHPEDASRFSRIPEGGGLRSCVSGKEFAKILLKIFTVELARHCDWQLNGBPPTMKTSPTV 480
DB 421 aphpedasrfsrifegggllrscvsgkefakillkiftvelarhcdwqlnbgpptomktsp 480
QY 481 YPVNDLPAFPTYEQGDI 497
DB 481 ypvndlparftfthgei 497

RESULT 5
AAW37733
ID AAW37733 standard; Protein; 492 AA.
XX
AC AAW37733;
XX
DT 07-JUL-1998 (first entry)

XX
DE Cytochrome zp450RAI protein.
XX
KW Retinoid regulated gene; cytochrome P450 gene; enzyme;
KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter.
XX
OS Danio rerio.
XX
PN W09749832-A2.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00488.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
PI Petkovich PM;
XX
DR WPI; 1998-077193/07.
DR N-PSDB; AAV09251.
XX

PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis

Example 1; Pages 53-55; 113pp; English.

CC This is the amino acid for cytochrome zp450RAI of the zebra fish.
CC Its expression is dependent on the presence of retinoic acid (RA).
CC The retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RAI is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen using
CC promoters and nucleotide sequences to identify drugs which are useful
CC for reducing the catabolism of RA.

Sequence 492 AA;

Query Match 66.4%; Score 1734; DB 19; Length 492;
Best Local Similarity 67.7%; Pred. No. 6e-159;
Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

QY 1 MGLPALLASALCTFVLPILLFLAALKLMDLYCVSSRDSCALPLPPGTMGFPFEGETLQM 60
DB 1 mglpallmvtflctlvlpvllflaavklwemlmirrvdpncrspippgtmgfpfgetlqi 60
QY 61 VLQRRKFLQMKRRKYGFITYKTHLFGPRTVRVMGADNVRRILLGEHRLVSVHWPASVRTL 120
DB 61 vlqrrkflmqmrrkygfiythlfgprtvrvmgadnvrrillgghklysvqwpasvrtll 120
QY 121 GAGCLSNLHDSHKKRKVKVIMQAFSREALQCYLVLYAEVSSCLEQWL---SCGERGLV 177
DB 121 gsdltisnvghyqhnkkkaimafsrdaehyipviqevksaigewlqkdscl---vly 176
QY 178 YPEVKRLMFRIAMRILGCEPBPAGGGEDEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGV 237
DB 178 ypevkrmfriamrillgcebpagggedeqolveafeemtrnlfslpidvpfsglyrgl 234
QY 238 KARNLIHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWEMGERLDMQALKOSTEL 297
DB 238 rarnfihskleenlrkkidq-ddnenegkykdalqllienstrsdepfslqamkaatei 293
QY 298 LFGHETTASAATSLITYLGLYPHVLOKVVREIRKSKGLCKSNODNKLDMETLEQLKYIG 357
DB 298 lfghettasaatslitylglyphvlokvvreelkskglcksnodnklhmeileqlkytg 353


```

QY      358 CVIKETLRINPPVPGCFRVALKTEFLNGYOIPKGWNVIYSICDTHDVADIFTNKEEFNP 417
         |||||||:|||||||:|||||||:|||||||:|||||||:|||||||:|||||||:
Db      354 cviketlrinppvpgrfvalktfelngyqipkgwnvysicdthdvadvfpnkeefgpe 413

QY      418 REIVPHPEDASRFSEFPGGLRSCVGKEFAKILKIETVELARHCDWQLNGPPTMKT 477
         ||: |||:||||| | |||||:|||| || |:| | |||||
Db      414 rfmkskledgsrfnylpgfgsrmcvgkfakvlkflvelqtghcnwilsngpptmktg 473

QY      478 PTVPVDNLPAFTYE 493
         ||:||||||| :||:
Db      474 ptlypdnlptkftsy 489

RESULT 6
AAW44159
ID      AAW44159 standard; Protein; 492 AA.
XX
AC      AAW44159;
XX
DT      22-JUN-1998 (first entry)
XX
DE      zebrafish retinoid metabolising protein zp450RAI.
XX
KW      Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW      retinoic acid; zebrafish; inhibitor; antibody; cancer;
KW      actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW      non-small cell lung carcinoma; basal cell carcinoma;
KW      acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW      ichthyosis; therapy; diagnosis; screening.
XX
OS      Danio rerio.
XX
PN      W09749815-A1.
XX
PD      31-DEC-1997.
XX
PE      23-JUN-1997; 97WO-CA00440.
XX
PR      01-OCT-1996; 96US-0724466.
PR      21-JUN-1996; 96US-0667546.
XX
PA      (TOOH ) UNIV QUEENS KINGSTON.
XX
PI      Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR      WPI; 1998-077178/07.
DR      N-PSDB; AAV12203.
XX
PT      Retinoid metabolising protein - useful to develop products to treat,
PT      e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT      ichthyosis
XX
PS      Claim 1; Page 50-51; 110pp; English.
XX
XX
XX      This protein comprises a novel zebrafish retinoid metabolising
CC      protein, designated ZP450RAI. Its amino acid sequence was deduced
CC      from a cDNA clone (see AAV12203) isolated from a 6-18 hr embryo
CC      library. It includes a haem-binding motif characteristic of
CC      cytochrome P450 proteins. ZP450RAI is a retinoid oxidase that has
CC      the ability to hydroxylate retinoic acid at the 4 position of the
CC      beta-ionone ring, and is inducible in epithelial cells exposed to
CC      retinoic acid. Zebrafish, human and mouse P450RAIs (see AAW44159-61)
CC      are claimed. They can be expressed in host cells and used to
CC      metabolize retinoic acid in an organism or cell, in drug screening,
CC      and to raise antibodies useful for inhibiting retinoic acid
CC      hydroxylation for the treatment of cancer, actinic keratosis, oral
CC      leukoplakia, secondary tumours of the head and/or neck, non-small
CC      cell lung carcinomas, basal cell carcinomas, acute promyelocytic
CC      leukaemia, skin cancer, and premalignancy associated with actinic
CC      keratosis, acne, psoriasis and/or ichthyosis.
XX
SQ      Sequence 492 AA;
```

Query Match	66.4%;	Score 1734;	DB 19;	Length 492;
Best Local Similarity	67.7%;	Pred. No. 6e-159;		
Matches 336;	Conservative 61;	Mismatches 89;	Indels 10;	Gaps 4;
QY	1	MGLPALLASALCTFEVPELLLFALAKLMDLYCVSSRDRSCALPLPGTMGPFPEFGETLOM	60	
Db	1	mglYtlmvtflctivlpvllflaavklwemlmirrvdpncrsplppgtmgjpfigetlqi	60	
QY	61	VLQRRKFLQMKRRKYGFTYKTHLFGRPTRVMGADNVRRIILGEHRLVSVHWPASVRTIL	120	
Db	61	ilgrkflrmkrqkygclykthlfgnptlvrmgadnvrqllgenklysvqwpasvrtll	120	
QY	121	GAGCLSNLHDSHKKQKKVYMQAFSREALQCYVLVIAEEVSSCLEQWL--SCGERGLV	177	
Db	121	gsdtlsnvhgvqhkknkkaimrafsrdalehyipvigevksaigewlqkdscc---vllv	176	
QY	178	YPEVKRLMFRIAMRILIGCEPAGGGEGDEOOLVEAFEEMTRNLFSLPIDVPFSGLYRGV	237	
Db	177	ypemkklmfriamrillgfepeq--ktdegelveafeemiknlfsipidvpfsglyrgl	234	
QY	238	KARNLIHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWGERLDMQALKOSSTEL	297	
Db	235	rarnfinskienirrkkiqd--dnenegkykdaqllliensrdsdeplsqamkeatel	293	
QY	298	LFEGHETTASATSLITYLGLYPHYLOKVREEIKSKGLLCKSNQDNKIDMETLEQLKYIG	357	
Db	294	lfqghettaatastslvmflglntevavqvreevqekvemgmtytpgkqlsmelldqklytg	353	
QY	358	CVIKETLRLNPVPVPGGFRVALKTFELNGYQIPKGWNVYISICDTHDVADIFTNKEEFNP	417	
Db	354	cviketrlrlnppvpvgfrrvalktfelngyqipkgwnviysicdthdvadvfpnkeefqpe	413	
QY	418	RFIVPHDEDASRESFIPEGGLRSCVGRFEAKILLKIFTVELARHCDMQLINGPPTMTKS	477	
Db	414	rflmskglegdsrfnyipfgggsrmcvgketaavlklkiflivelqtghcnwlinsgppmtkty	473	
QY	478	PTVVPVDNLPARFTYF 493		
Db	474	ptliypvdnlpkrftsy 489		

RESULT	7	
AAB12489		
ID	AAB12489	standard; Protein; 216 AA.
XX		
AC	AAB12489;	
XX		
DT	31-OCT-2000	(first entry)
XX		
DE	Human PSEC64	protein sequence SEQ ID NO:2.
XX		
KW	Human; PSEC64;	neuron growth; nerve disease.
XX		
OS	Homo sapiens.	
XX		
PN	JP2000152790-A.	
XX		
PD	06-JUN-2000.	
XX		
PF	19-NOV-1998;	98JP-0329989.
XX		
PR	19-NOV-1998;	98JP-0329989.
XX		
PA	(HERI-)	HERIKUSU KENKYUSHO KK.
XX		
DR	WPI; 2000-468126/41.	
DR	N-PSDB; AAA60752.	
XX		
PT	A protein related to the growth of neuron and a gene encoding said	
XX	protein -	

PS Claim 1; Page 10; 13pp; Japanese.

CC The present sequence represents a human protein, designated PSEC64, which
CC is related to neuron growth. The PSEC64 protein and its gene can be used
CC for the development of a preventive agent for use in the treatment of
CC diseases in which nerves are involved.

XX Sequence 216 AA;

Query Match 40.3%; Score 1052; DB 21; Length 216;
Best Local Similarity 93.5%; Pred. No. 1.5e-93;
Matches 202; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MGLPALIASALCTFVLPPLLFALAKLMDLYCVSSRDRSCALPPPGTMGFPFFGETLQM 60
Db 1 mglpallasalctfvlplllflaalklwdlycvsgdrscalp1ppgtmgfiffgetlqm 60
QY 61 VLQRRKFLQMKRRKYGFYIKTHTLFGRPVTVRMGADNVRRIILGHEHRLVSVHWPASVRTIL 120
Db 61 v1qrrkflqmkrrkygfyytkthlfgrrptvrvmgadnvrriilghehrlvsvhwpasvrtil 120
QY 121 GAGCLISNLHDSHKKQRRKVMQAFSREALQCYVLVIAEEVSSCLEQWLSCGERGLLVPE 180
Db 121 gsgclsnlhdsshkqrkvmratsrealcycvpvitteevgsleqwlscgergllype 180
QY 181 VKRMERIAMRIILGCEPAGCGEDEQQLVEAFEE 216
Db 181 vkrmftriamrillgcepq1agdgdsseqqlveafee 216

RESULT 8

AAG20783
ID AAG20783 standard; Protein; 468 AA.

XX AAG20783;
DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 23102.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.

PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151308.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159295.
PR	13-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160778.
PR	21-OCT-1999;	99US-0160814.

PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	18.1%;	Score 474;	DB 21;	Length 468;
Best Local Similarity	27.38;	Pred. No. 3e-37;		
Matches 134;	Conservative	93;	Mismatches 214;	Indels 50;
				Gaps 12;

```

QY      15 VLPLLLFLAALKMLNDLYCVSSRRDSCALBPPGTMGFPEFGETLQMLQÖR-RKFLQMKRR 73
      1 : : : :      |      | : : : :      |      | : :
Db      8 vvpilllcllllv---rvivskkkknstrgklppsgmawpylgetelqlysqnpuvfetskqk 64

```

QY 74 KYGFITYKTHLFGRPITVRVMGADNVRILLGEHRLVSVHWMPASVPTILGAGCLSNLDSSH 133
::|||::|||::|||::|||::|||
Db 65 rygeifktrilgyepcvmlaspeaarfvltahnmfkptyriskreklipsalffhgqdyh 124

QY 134 KQRKVVIMQAFSREALQCYVLVIAEVSQCLEQWLSGGERGLVYPEVKRLMFRIAMRIL 193
::: :| : : : :| : : : :| :| : : : :
Db 125 shirkIvqssfyetirklipdiehialssIqswanmpi--vscyqemkfkafdvglai 182

```

QY      194  LGCEPGAGGEDEQLVEAFEEINTRLF-----SLPIDVPEGLYRGVKARNLTHA 245
      |      |      |      |      |      |      |      |      |      |
Db      183  fq-----hlessykeilkhnynivdkqynsfpmslpqtshykalmarqkqlt 229

```

```

QY      246 RIEENT--RAKIRRLQATEPDGGCKDALQZLLIEHSWGERGERLDMQALKQSSTELLFGGHE 303
          : | | | : | | | | | | : | | | : : : : | | | :
Db      230 lvsEICerrekralqt-----dfIghllfnknekrvltqgeIadniiIgvlfaaqd 281

```

QY 304 TTASATSLITYLGLYPHYLVQKVRREIKS--KGLICKSNQDNK-IDMETLEQLKYIGCVI 360
||||| : : ||| | ||: | : | : : : : | :
Db 282 ttascsltwilkyl---hddqklleavkaeqkaiyeensrekkplwtqrtnmplthkvi 337

QY 361 KETLRINPPVPGGFRAALKTFELNGYQIPKGNVYISICDTHDVADIFTNKEEFNPDREI 420
|::|: : | | | | | : : | : | | | | |
Db 338 vestlmasiisftfreaydvdekvqylipkwkymplfnrlnhmkpysnpevfdsprfe 397

```

QY      421 V-PHPEDASRFSFIPEGGCLRSCVCGKEFAKILKIFTELARHCDMOLLNGRPPTMKTSP 479
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db      398 vnohon-----tfmofasvvhacpapelaklqilflhlnvsnfirwevbxqekqiydspf 452

```

QY	480	VVPYDNLPAF	490
		:	
nh	453	njpnnqjpatf	463

RESULT	9
AAB07921	
ID	AAB07921 standard; Protein; 513 AA.

DT	14-NOV-2000 (first entry)
XX	
DE	A cytochrome P450 enzyme designated DMF4.

XX
KW DWF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
KW plant phenotype; cell elongation.

Arabidopsis sp.

AA
PN WO200047715-A2.

[illegible]

PD 17-AUG-2000.
XX
XX 11-FEB-2000; 2000WO-US03820.
PF
XX 11-FEB-1999; 99US-0119657.
PR 11-FEB-1999; 99US-0119658.
XX
PA (ARIZ-) ARIZONA BOARD OF REGENTS.
XX
PI Azpiroz R, Choe S, Feldmann KA;
XX
XX WPI; 2000-549142/50.
DR N-PSDB; AAA59599.
DR
XX
PT New isolated dwf4 polynucleotide useful for altering the phenotype of
PT plants, for diagnostic assays and in the production of antibodies -
XX
PS Claim 50; Fig 11; 113pp; English.
XX
XX The present sequence represents a DWF4 polypeptide. The polypeptide is a
CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
CC brassinosteroids. Specifically, it mediates multiple
CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
CC polynucleotide is used for altering the phenotype of a plant. DWF4
CC plants display a dramatic reduction in the length of different organs,
CC and this size reduction is attributable to a defect in cell elongation.
CC The DWF4 polynucleotides and polypeptides can be used in diagnostic
CC assays and to generate antibodies, which can be used to produce
CC immunogenic compositions.
CC
XX
SQ Sequence 513 AA;

Query Match 17.6%; Score 461; DB 21; Length 513;
Best Local Similarity 28.4%; Pred. No. 6.2e-36;
Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;
QY 3 LPALIASALCTFVLPPLLFLAALKIMDLVCSRRDRSCALPLPGTMGFEEGETL----- 58
Db 10 lpllllpsl----lslllflilk-----rrnrktrfnlpgksqwpflgetigyik 57
QY 59 -QWVLORRKFLQMKRRKYGFIYKTHLFGRPTRVMGADNVRRLILGEHRLVSVHWPASVR 117
Db 58 pytatlgdmqhvskygkiyrnlfgeptlvsadaglnrfilqnegrlfecysprsig 117
QY 118 TILGAGCLSNLHDSHKQKKVIMQAFSREALQCYLVV-IAEVSCLLEQWLSCGERGLL 176
Db 118 gllgkwsmlvlvgdmhrmsislnflsharlrtlllkdverhtlflvldswqg--nsifs 175
QY 177 VPEYKRLMERIAMRILGCEPAPAGGEDEQQLVEAFEEEMTRNLESLPIDVPFSGLYRG 236
Db 176 aqdeakkftfnlmakhimsmdpge-----eeteglkeyvtfmkgvvsaplnlpgtayhka 231
QY 237 VKARNULIHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWE-----RGERLDMQ- 288
Db 232 lqsratilkfierknee--rkldikeed---qeeevktedeaemsksdhvrkqrtdddl 286
QY 289 ---ALKQS--STE-----LLFGHETTASATSLITYLGLYPHVLQKVREIISKGL 335
Db 287 lgvwlkhsnlstegldlslsllfaghetssvaialaiflqacpkaaveelreehleiar 346
QY 336 LCKSNQDNKLDMETLEQLKYIGCVIKETLRLNRPVPGGFVALKTFELNGYQIPKGMNVI 395
Db 347 akkelgeselnwddykkmddfqcvinetlrlgnvvrflhrkalkdvrykgydipsgwkv1 406
QY 396 YSICDTHDVADIFTNKEEFNDRFIVRPHEDASRFS-----FIPEGGGLRSCVCKE 446
Db 407 pvisavhldnsrydqpnlfnpwrwqgqngasssgsfstgwgnympfggpprlcagse 466
QY 447 FAKILLKIFTVELARHCDWQLNGRPRTMKTSPTVYPPVDNLP 487
Db 467 laklemavfihlhlvklfnwelaee-----ddqpfafpfvdfp 502

RESULT 10
AAG46490
ID AAG46490 standard; Protein; 461 AA.
XX
AC AAG46490;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58494.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137724.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 17.4%; Score 455.5; DB 21; Length 461;
Best Local Similarity 28.2%; Pred. No. 1.8e-35;
Matches 137; Conservative 91; Mismatches 211; Indels 47; Gaps 13;

QY 1 MGLPALASACTFVLPLLLFLAALKLMDLYCVSSRDRSCALPPEPGTMGPFPGETLQM 60
Db 4 mgl111ivs-1c-----sallrw-----ngmrytknglppgtmgwplfgettef 46
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGRPYVYMGADNVRRIILGHRILVSVHWPASVRTIL 120
Db 47 lkqpnfmtrnqrlrygsffkshllgcptlismdsevnryllkneskglvpyqpsmdl1 106
QY 121 GAGCLSNLHDSHQRKKVIMQAFSREALQCYLVLIAEE-VSSCLEQWLSCGERGLL-VY 178
Db 107 gtcmaavhgshtlmrgslslssttmrchilpkvdhfmrsyldqw---nelevldiq 163

PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 17.4%; Score 455.5; DB 21; Length 462;
Best Local Similarity 28.2%; Pred. No. 1.8e-35;
Matches 137; Conservative 91; Mismatches 211; Indels 47; Gaps 13;

QY 1 MGLPALASALCTFVLPDLLFLALAKLMDLYCVSSRDRSCALPLPPTMGFPFGETLQM 60
Db 5 mgl1111ivs-1c-----sall1rw-----ngmrytknglppgtmgwplfgettef 47
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGRPTRVVMGADNVRRILLGEHRLVSVHWPASVFTIL 120
Db 48 lkgqpnfmrnqrlrygsffkshllycptlismdsevnryllkneskylvpypqsmldl1 107
QY 121 GAGCLSNLHDSHKKQKKVIMQAFSREALQCYLVIAEE-VSSCLEQWLSCGERGLI-VY 178
Db 108 gtcmaavhygsshrmlrgslslsistmmrdhllpkvdfmrSYldqw---nelevidiq 164
QY 179 PEVKRLMFRAMRILLGCEPAGGGEDEQOLVEAFEEEMTRNLFSLPIDVPFSGLRGVK 238
Db 165 dtkkhmaflssltqdiagnlrkpf-----veefktafklvvgtlsvpidlpgtlnyrqld 219
QY 239 ARNLIHARIEENIRAKIRRLQATEPDDG--CKDALQLLIEHSWGERLDMQALKOSSTE 296
Db 220 arn-----nidrllreImgeridsgetftdmlylmkqegnrypltd-eairdqvt 270
QY 297 LLEGGHETTSAAATSLITYLGLYPHYLOKVREEIKSKGLCKSNODNKLDMETLEQLKYI 356
Db 271 llysgyetvstsmalkylldhpkalqelrae--hlafrerkrqdeplgledvksmkft 328
QY 357 GCVIKETLRILNPVPGGFRVALKTFEELNGYQIPKGNVYISICDTHDVADIFTNKEEFNP 416
Db 329 raviyetstrlatvngvrlktrdldengylilpkgyrvytreinydanlyedplifnp 388
QY 417 DRFTVPHPEDASRFSFIPEGGGLRSCVQKEFAKILKIFTVELARHCDWQLNGPPTMKT 476
Db 389 wrwmkksle--sqnscfvfggtrlcpqkelgiveissflhyfvrtryweeligdelm-- 444
QY 477 SPTVYP 482
Db 445 ---vfp 447

RESULT 12
AAG46491
ID AAG46491 standard; Protein; 465 AA.

XX AAG46491;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 58495.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58495.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX 06-SEP-2000.
XX PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX

PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	04-MAY-1999;	99US-0132407.
PR	05-MAY-1999;	99US-0132484.
PR	06-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132487.
PR	11-MAY-1999;	99US-0132863.
PR	14-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137752.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.

PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147202.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.

99US-01435421	99US-01436224	99US-01440051	99US-01440851	99US-01440861	99US-01443225	99US-01443321	99US-01443325	99US-01443331	99US-01443333	99US-01443334	99US-01443335	99US-01443352	99US-01443521	99US-01448841	99US-01450861	99US-01450881	99US-01450887	99US-01450891	99US-01451082	99US-01451145	99US-01452181	99US-01452224	99US-01452276	99US-01452913	99US-01459119	99US-01459119	99US-01459511	99US-01463861	99US-01463881	99US-01463891	99US-01470381	99US-01472024	99US-01473021	99US-01471192	99US-01472260	99US-01473303	99US-01474116	99US-01474931	99US-01479351	99US-01481711	99US-01483119	99US-01485651	99US-01486841	99US-01493368	99US-01491751	99US-01494261	99US-01497221	99US-01497231	99US-01499291	99US-01499302	99US-01499303	99US-01505661	99US-01508841	99US-01510651	99US-01510661	99US-01510801	99US-01513031	99US-01514381	99US-01519301	99US-01523631	99US-01530701	99US-01537581	99US-01540181	99US-01540391	99US-01547791	99US-01551391	99US-01554861	99US-01556581	99US-01564581	99US-01565961
---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------

PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 17.4%; Score 455.5; DB 21; Length 465;
Best Local Similarity 28.2%; Pred. No. 1.8e-35;
Matches 137; Conservative 91; Mismatches 211; Indels 47; Gaps 13;

QY	1	MGLPALLASALCTFVLPULLFLAALKLMDLYCVSSRDRSCALPLPPGTMGFPFEGETLQM	60
Db	8	mgllllivs-lc-----sallrw-----ngmrytknglppgtmgwplfgettef	50
QY	61	VLQRRKFLQMKRRKYGFITYKTHLFGRPYRVWGADNVRRIILGEHRLVSVHWPAVRTIL	120
Db	51	lkqgnfmnrqrlxygsfiksllgcptllismdsevnryllkneskqjvpgypqsmldll	110
QY	121	GAGCLSNLHDSHKKQKKVIMQAFSREALQCYVLVIAEE-VSSCLEQWLSCGERGLL-VY	178
Db	111	gtcmaavhgsshrlnrsgllslistmmrthlplkvdhfmsryldqw---nelevidiq	167
QY	179	PEVKRLMFRIAMRILLGCEPGPAGGEDQQLVFAFEEMTRNLFSLPIDVPFSGLYRGVK	238
Db	168	dktkmaflssltqtaglnlrpf----veefktaffklvgtlsvpidpgtntyrcgiq	222
QY	239	ARNLIHARIEENIRAKTRRLQATEPDDG--CKDALQLLIEHSWGERLDMQALKQOSTE	296
Db	223	arn-----nidrlrlrelmqerrdsgetftdmlylmkkegnrypltd-eelrdqvt	273
QY	297	LEFGHETTASATSLITYLGLYPHVLQKVREETKSGLLCKSNODNKLDMETLEQLKYI	356
Db	274	llysgyetvststmmalkylhdpkalgelrae--hlafrerkrqdepljledvksmkft	331
QY	357	GCVIKETLRLNPPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADIFTNKEFNP	416
Db	332	ravlyetsrlativngvjlkrtdrleingyllpkgwriyvvtreinydanlyedplifnp	391
QY	417	DREIVPHPEDASRFSFIPGGGLRSCVSGKEFAKILKIFTVELARHCDWQLNGPPTMKT	476
Db	392	wymkksle--sqnscfvgggtrlcpkgelgiveissflhyfvtryrweeiyggdelm--	447
QY	477	SPTVYP 482	
		l:l	

Db	448	---vfp 450
RESULT 13		
AAW27153	standard; Protein; 472 AA.	
ID	AAW27153	
XX	AAW27153;	
AC	AAW27153;	
XX	14-APR-1998 (first entry)	
DE	Arabidopsis thaliana cytochrome P450-type hydroxylase.	
XX	Cytochrome P450-type hydroxylase; identification; brassinosteroid;	
KW	brassinosteroid inhibitor; modified plant; recombinant production;	
KW	teasterone.	
XX	Arabidopsis thaliana.	
OS	Arabidopsis thaliana.	
XX	WO9735986-A1.	
PN	WO9735986-A1.	
XX	02-OCT-1997.	
PD	02-OCT-1997.	
XX	27-MAR-1997; 97WO-EP01586.	
PF	27-MAR-1997; 97WO-EP01586.	
XX	27-MAR-1996; 96US-0622166.	
PR	27-MAR-1996; 96US-0622166.	
XX	(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.	
PA	(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.	
XX	Altman T, Koncz C, Mathur J, Szekeeres MA;	
PI	Altman T, Koncz C, Mathur J, Szekeeres MA;	
XX	WPI; 1997-489649/45.	
DR	N-PSDB; AAT85306, AAT85307.	
DR	N-PSDB; AAT85306, AAT85307.	
XX		
PT	New isolated plant cytochrome P450-type hydroxylase gene - used to	
PT	identify substances acting as brassino-steroid(s) or brassinosteroid	
PT	inhibitors for the production of modified plants	
XX		
PS	Claim 1; Pages 44-46; 77pp; English.	
XX		
CC	The present sequence is Arabidopsis thaliana cytochrome	
CC	P450-type hydroxylase. The hydroxylase can be used to identify	
CC	brassinosteroids or brassinosteroid inhibitors, useful to produce	
CC	plants with modified physiological and/or phenotypic	
CC	characteristics. The modified plants may show, e.g. stimulation of	
CC	growth, increased cell elongation, increased wood production, an	
CC	accelerated seed germination at low temperatures, an increase in	
CC	dry weight, repressed anthocyanin production during growth in light	
CC	and/or inhibited de-etiolation which is induced, e.g. by cytokinin,	
CC	in the dark or an increase in stress tolerance. The hydroxylase or	
CC	its coding sequence can also be used for the recombinant production	
CC	of compounds, e.g. teasterone.	
XX		
SQ	Sequence 472 AA;	

Query Match 17.2%; Score 450; DB 18; Length 472;
Best Local Similarity 26.6%; Pred. No. 6.3e-35;
Matches 134; Conservative 90; Mismatches 217; Indels 62; Gaps 15;

QY	6	LIASALCTFVLPULLFLAALKLMDLYCVSSRDRSCALPLPPGTMGFPFEGETLQMV----	61
Db	9	llssiaagflllll-----rrtryrmglppsglqlpligetfqlgayk	52
QY	62	-LQRRKFLQMKRRKYGFITYKTHLFGRPYRVWGADNVRRIILGEHRLVSVHWPAVRTIL	120
Db	53	tenpepfidervarygsvfmthlfgcptlfsadpetnrfvlnqegklfecsyasicnll	112
QY	121	GAGCLSNLHDSHKKQKKVIMQAFSREALQCYVLV-IAEEVSSCLEQWLSCGERGLAYP	179
Db	113	gkhslllmkyslhkrmsltsfansiikdhlmldidrlvrfnldswss----rvllme	168
QY	180	EVKRLMFRIAMRILLGCEPGPAGGEDQQLVFAFEEMTRNLFSLPIDVPFSGLYR-GVK	238

Db 169 eakkitfeltvklmsfdp-----gwseslirkeylllviegffslppl-fsttyrkaig 222
QY 239 ARNLIHARIEENTIRAKIRRLQATEPDGG--CKDALQLLI--EHSWERGERLDMQALKOSS 294
Db 223 arr----kvaaaltvvmkireeeegaerkkmlaallaaddgfsdeeydf-----1 272
QY 295 TELLFGGHETYSAAATSLITYLGLYPHVLQKYREERIKSKGLLCKSNQDNKIDMETLEQLK 354
Db 273 vallvaagyettstlmtlavkfltetplalaqlkee--hekiramksdsyslewsdykmp 330
QY 355 YIGCVIKETLRINPPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHVDIFTNKEEF 414
Db 331 ftqcvvnetlrvanligvfrirramtdvealkgykikgwkvfssfravhldpnhktdartf 390
QY 415 NPDRF---IYPHPEDASRFSFIPFGGGLRSCVGKEFAKILLKIFTVELARHCDWQLLNG 470
Db 391 npwrwgsnsvtctgpsnv----ftpfggprlcpgyelaravalsvflhrlvtgfsw----- 441
QY 471 PPTMKTSPFVYPVDNLPARFTYF 493
Db 442 vpaegdklvffprrttrtkrypif 464

RESULT 14

AAG44571
ID AAG44571 standard; Protein; 472 AA.

XX AAG44571;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55847.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 17.28; Score 450; DB 21; Length 472;
Best Local Similarity 26.68; Pred. No. 6.3e-35;
Matches 134; Conservative 90; Mismatches 217; Indels 62; Gaps 15;

QY 6 LLASALCTFVLPPLLFLAALKIMDLVCVSSRDRSCALPLPPTMGFPFPGETIQMV---- 61
Db 9 llsiaagfllll-----rrtryrmqjppsglglpligetqligayk 52
QY 62 -LORRKFLQMKRRKYGFIYKTHLEGRPTVRVMGADNVRRILLGEHRLVSVHWPASVRTIL 120
Db 53 tenpepfidervarygsvfmtlhlfgeptlfsadpetnrfvlnegkllfecsyasicnll 112
QY 121 GAGCLSNLHDSHKQKKVIMQAFSREALQCYLVV-IAEVSVCLEQWLSCGERGLLYP 179
Db 113 gkhslllmkyslhrkmsltsfanssikhlmldldrlvrfnldswss---rvllme 168
QY 180 EVKRLMFRIAMRILLGCEPFGGGEDEQQLVEAFEEEMTRNLFSLPIDVPFSGLYR-GVK 238
Db 169 eakktiteltvkqlmsfdp-----gwseslrkeyllviegffslplpl-fsttyrkaiaq 222
QY 239 ARNLIHARIEENIRAKIRRLQATEPDG--CKDALQLLI--EHSWGERLDMQALKOSS 294
Db 223 arr---kvaecalrvvmkrrreeeegaerkkdmlaallaaddgfsdeeivdf-----l 272
QY 295 TELLEGGHETTASAATSLITYLGLVPHYLQKVREEIKSKGLICKSNQDNKLDMEETLEQLK 354
Db 273 vallvagyetlstimtlavkflletplalaqlkee--hekiramksdsyslewsdyksmp 330
QY 355 YIGCVIKETRLRNPPVPGFRRVALKTFELNGYQIPKGMVNIYSICDTHDVADIFTKKEEF 414
Db 331 ftqcvnetlrvanilgvlrramdvelkgykirkpkywkvfssfravhldpnhfdaartf 390
QY 415 NPDRF---IVPHPEDASRFSFIDPFGGGLRSCVGKEFAKILKIFTELARHCDWQLING 470
Db 391 npwrwgsnsvtqpsnv----ftpfqg99prlcpyelarvalsvflhrlyvtgfsw----- 441
QY 471 PPTMKTSPTYXPVDNLPARFTYF 493
Db 442 vpaegdklvffpplrtrtqkryplf 464

RESULT 15
AAG45022
ID AAG45022 standard; Protein; 472 AA.
XX AAG45022;
AC AAG45022;
XX 18-OCT-2000 (first entry)
DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 56469.
DE Arabidopsis thaliana
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX EP1033405-A2.
PN EP1033405-A2.
XX 06-SEP-2000.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PF 25-FEB-1999; 99US-0121825.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.

PR	09-MAR-1999;	99US-0123548;
PR	23-MAR-1999;	99US-0125788;
PR	25-MAR-1999;	99US-0126264;
PR	29-MAR-1999;	99US-0126785;
PR	01-APR-1999;	99US-0127462;
PR	06-APR-1999;	99US-0128234;
PR	08-APR-1999;	99US-0128714;
PR	16-APR-1999;	99US-0129845;
PR	19-APR-1999;	99US-0130077;
PR	21-APR-1999;	99US-0130449;
PR	23-APR-1999;	99US-0130510;
PR	23-APR-1999;	99US-0130891;
PR	28-APR-1999;	99US-0131449;
PR	30-APR-1999;	99US-0132048;
PR	30-APR-1999;	99US-0132407;
PR	04-MAY-1999;	99US-0132484;
PR	05-MAY-1999;	99US-0132485;
PR	05-MAY-1999;	99US-0132486;
PR	05-MAY-1999;	99US-0132487;
PR	07-MAY-1999;	99US-0132863;
PR	11-MAY-1999;	99US-0134216;
PR	14-MAY-1999;	99US-0134218;
PR	14-MAY-1999;	99US-0134219;
PR	14-MAY-1999;	99US-0134221;
PR	14-MAY-1999;	99US-0134370;
PR	18-MAY-1999;	99US-0134768;
PR	19-MAY-1999;	99US-0134941;
PR	20-MAY-1999;	99US-0135124;
PR	21-MAY-1999;	99US-0135353;
PR	24-MAY-1999;	99US-0135629;
PR	25-MAY-1999;	99US-0136021;
PR	27-MAY-1999;	99US-0136392;
PR	28-MAY-1999;	99US-0136782;
PR	01-JUN-1999;	99US-0137222;
PR	03-JUN-1999;	99US-0137528;
PR	04-JUN-1999;	99US-0137502;
PR	07-JUN-1999;	99US-0137724;
PR	08-JUN-1999;	99US-0138094;
PR	10-JUN-1999;	99US-0138540;
PR	10-JUN-1999;	99US-0138847;
PR	14-JUN-1999;	99US-0139119;
PR	16-JUN-1999;	99US-0139452;
PR	16-JUN-1999;	99US-0139453;
PR	17-JUN-1999;	99US-0139492;
PR	18-JUN-1999;	99US-0139454;
PR	18-JUN-1999;	99US-0139455;
PR	18-JUN-1999;	99US-0139456;
PR	18-JUN-1999;	99US-0139457;
PR	18-JUN-1999;	99US-0139458;
PR	18-JUN-1999;	99US-0139459;
PR	18-JUN-1999;	99US-0139460;
PR	18-JUN-1999;	99US-0139461;
PR	18-JUN-1999;	99US-0139462;
PR	18-JUN-1999;	99US-0139463;
PR	18-JUN-1999;	99US-0139750;
PR	18-JUN-1999;	99US-0139763;
PR	21-JUN-1999;	99US-0139817;
PR	22-JUN-1999;	99US-0139899;
PR	23-JUN-1999;	99US-0140353;
PR	23-JUN-1999;	99US-0140354;
PR	24-JUN-1999;	99US-0140695;
PR	28-JUN-1999;	99US-0140823;
PR	29-JUN-1999;	99US-0140921;
PR	30-JUN-1999;	99US-0141287;
PR	01-JUL-1999;	99US-0141842;
PR	01-JUL-1999;	99US-0142055;
PR	02-JUL-1999;	99US-0142054;
PR	06-JUL-1999;	99US-0142390;
PR	08-JUL-1999;	99US-0142803;
PR	09-JUL-1999;	99US-0142920;
PR	12-JUL-1999;	99US-0142977;
PR	13-JUL-1999;	99US-0143542;
PR	14-JUL-1999;	99US-0143624;

PR	15-JUL-1999;	99US-0144005;
PR	16-JUL-1999;	99US-0144006;
PR	16-JUL-1999;	99US-0144085;
PR	19-JUL-1999;	99US-0144325;
PR	19-JUL-1999;	99US-0144331;
PR	19-JUL-1999;	99US-0144332;
PR	19-JUL-1999;	99US-0144333;
PR	19-JUL-1999;	99US-0144334;
PR	19-JUL-1999;	99US-0144335;
PR	20-JUL-1999;	99US-0144352;
PR	20-JUL-1999;	99US-0144632;
PR	20-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0145086;
PR	21-JUL-1999;	99US-0145088;
PR	22-JUL-1999;	99US-0145085;
PR	22-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145089;
PR	22-JUL-1999;	99US-0145192;
PR	23-JUL-1999;	99US-0145145;
PR	23-JUL-1999;	99US-0145224;
PR	23-JUL-1999;	99US-0145226;
PR	26-JUL-1999;	99US-0145276;
PR	27-JUL-1999;	99US-0145913;
PR	27-JUL-1999;	99US-0145918;
PR	27-JUL-1999;	99US-0145919;
PR	28-JUL-1999;	99US-0145951;
PR	02-AUG-1999;	99US-0146386;
PR	02-AUG-1999;	99US-0146388;
PR	02-AUG-1999;	99US-0146389;
PR	03-AUG-1999;	99US-0147038;
PR	04-AUG-1999;	99US-0147204;
PR	04-AUG-1999;	99US-0147302;
PR	05-AUG-1999;	99US-0147192;
PR	05-AUG-1999;	99US-0147260;
PR	06-AUG-1999;	99US-0147303;
PR	06-AUG-1999;	99US-0147416;
PR	09-AUG-1999;	99US-0147493;
PR	09-AUG-1999;	99US-0147935;
PR	10-AUG-1999;	99US-0148171;
PR	11-AUG-1999;	99US-0148319;
PR	12-AUG-1999;	99US-0148341;
PR	13-AUG-1999;	99US-0148565;
PR	13-AUG-1999;	99US-0148684;
PR	16-AUG-1999;	99US-0149368;
PR	17-AUG-1999;	99US-0149175;
PR	18-AUG-1999;	99US-0149426;
PR	20-AUG-1999;	99US-0149722;
PR	20-AUG-1999;	99US-0149723;
PR	20-AUG-1999;	99US-0149929;
PR	23-AUG-1999;	99US-0149902;
PR	23-AUG-1999;	99US-0149930;
PR	25-AUG-1999;	99US-0150586;
PR	26-AUG-1999;	99US-0150884;
PR	27-AUG-1999;	99US-0151065;
PR	27-AUG-1999;	99US-0151066;
PR	27-AUG-1999;	99US-0151080;
PR	30-AUG-1999;	99US-0151303;
PR	31-AUG-1999;	99US-0151348;
PR	01-SEP-1999;	99US-0151930;
PR	07-SEP-1999;	99US-0152363;
PR	10-SEP-1999;	99US-0153070;
PR	13-SEP-1999;	99US-0153758;
PR	15-SEP-1999;	99US-0154018;
PR	16-SEP-1999;	99US-0154039;
PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-0155139;
PR	23-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157117;
PR	05-OCT-1999;	99US-0157753;

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 17.2%; Score 450; DB 21; Length 472;
Best Local Similarity 26.6%; Pred. No. 6.3e-35;
Matches 134; Conservative 90; Mismatches 217; Indels 62; Gaps 15;

QY 6 LLASALCTFVLPILLELALIKLMDLYCVSSRDRSCALPLPPGTMGFPFGETLQMV---- 61
Db 9 llsiaagfllll-----rrtryrmglppsglpligetfqligayk 52

QY 62 -LQRRKFLQMKRRKYGFYKTHLFGPTRYVMGADNVRRIILGHEHRLVSVHWPASVRTIL 120
Db 53 tenpepfidervarygsvmthlfgpeplfsadpetnrflqneqklfeesypasicnll 112

QY 121 GAGCLSNLHDSHKKRVIMQAFSREALQCYVIV-IAEVSSCLEQWLSCGERGLLYP 179
Db 113 gkhslllmkyslkrmhslmsfanssilkdhlmldrlvrfnldsws----rvllme 168

QY 180 EVKRLMFRIAMRILGCEPAGGEDEQOLVEAFEEMTRNLFSLPIDVPFSGLYR-GVK 238
Db 169 eakkitfeltvkqlmsfdp----gewseslrkeyllviegffslplpl-fstlyrkaig 222

QY 239 ARNLIHARIEENIRAKIRRLQATEPDG--CKDALQLLI--EHSWGERLDMQALKOSS 294
Db 223 arr----kvaeaaltvymkrreeeegeaerkkmdlaallaaddgfsdeelvdf-----1 272

QY 295 TELLFGHETASATSLITYLGLYPHVLOKVREITKSKGLLCKSNQDNKIDMETLEQLK 354
Db 273 vallvagyetstlmtlavkfletplalaqlkee--hekiramksdsyslewsdyksmp 330

QY 355 YIGCVIKETLRNLNPPYGGFRVALKTFELNGYQIPKGNVIYSICDTHDVADIFTNKEEF 414
Db 331 ftqcvvnetlrvanilgyfrramtvelkgyktpkgwkvfssfravhdpnhfkdarlf 390

QY 415 NPDRE----IVPHPEDASRFSFIPEGGLRSCVGEKFAKILKIFTVELARHCDWQLNG 470
Db 391 npwrwqsnsvltgpsnv----ftpfggprlcpgyelarvalsvflhrllvtgfsw----- 441

QY 471 PPTMKTSPTVYPVDNLPARFTYF 493
Db 442 vpaeqdklvffpttrtqtqkryplf 464

Search completed: November 6, 2001, 13:25:27
Job time: 12655 sec

OM of: US-09-668-482-4 to: EST:* out_format : pfs
Date: Nov 6, 2001 2:26 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed+p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US09668482/runat_05112001_133353_10873/app_query.fasta_1.1680
-DB=EST -QEMT=fastap -SUFFIX=first -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEX=7.000 -START=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTEMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09668482@cgn1_1_5307 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-668-482-4
Query length: 497
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 2793.150000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_est29:AL539668	+ 1431.50	2868.54	1.2e-150	974	AL539668 LTI_FL013_FBrn1
gb_est29:AL532445	+ 1376.00	2757.18	1.9e-144	893	AL532445 LTI_NFL001_NE
gb_est29:AL532444	- 1313.00	2627.94	2.9e-137	1028	AL532444 LTI_NFL001_N
gb_est48:AM513600	+ 1045.00	2091.25	2.3e-107	618	AM513600 x047b06.x1 NCI_CGAP_U
gb_est29:AL539667	- 1024.00	2045.01	8.7e-105	847	AL539667 AL539667 LTI_FL013_FBrn1
gb_est72:BE236243	+ 881.00	1760.81	5.8e-89	537	BE236243 143959 MARC IPiG Sus s
gb_est87:BF385034	+ 852.00	1702.05	1.1e-85	539	BF385034 602045632F1 NCI_CGAP_I
gb_est89:BF533769	+ 852.00	1701.51	1.2e-85	564	BF533769 602073982F1 NCI_CGAP_I
gb_est71:BE189825	+ 840.00	1674.73	3.6e-84	696	BE189825 db61c05.y1 Wellcome C
gb_est4:AA239785	+ 823.00	1644.84	1.7e-82	474	AA239785 mx80a03.r1 Soares mous
gb_est85:BF236872	+ 769.00	1533.84	2.6e-76	545	BF236872 602027980F1 NCI_CGAP_I
gb_est82:BF055367	- 763.00	1519.69	1.6e-75	646	BF055367 7j78f12.x1 Soares NSF
gb_est43:AM174347	+ 751.50	1495.12	3.7e-74	720	AM174347 f142a06.y1 Sugano Kawa
gb_est51:AM765767	+ 731.50	1455.49	5.9e-72	669	AM765767 da77a02.y1 Harland sta
gb_est23:AI667038	+ 691.50	1375.01	1.8e-67	640	AI667038 fc24h03.y1 Zebrafish W
gb_est107:R51129	+ 680.50	1355.44	2.2e-66	509	R51129 y971b05.r1 Soares infant
gb_est43:AM128823	+ 673.50	1339.91	2.6e-65	571	BS15142 da60f10.y1 Wellcome C
gb_est102:BG578446	+ 642.50	1276.61	5.5e-62	597	BG578446 de99c02.y1 Wellcome C
gb_est99:BG364937	+ 642.00	1277.18	5.1e-62	522	BG364937 dc93f02.y1 NICHHD XGC C
gb_est82:BF047649	+ 624.00	1241.33	5.0e-60	496	BF047649 dc80h02.y1 NICHHD XGC C
gb_est87:BF426259	+ 617.00	1226.78	3.3e-59	512	BF426259 df69e06.y1 Xenopus lae
gb_est49:AM595907	+ 612.00	1215.54	1.4e-58	563	AM595907 fa29g11.y2 Xenla 13LiC
gb_est43:AM171465	- 588.00	1163.32	1.1e-55	765	AM171465 f142a06.x1 Sugano Kawa
gb_est44:AM244588	- 582.00	1159.90	1.7e-55	365	AM244588 BR_END09D04ps2 Bain Rd
gb_est14:AI004222	- 579.00	1149.03	6.7e-55	532	AI004222 ou03e01.x1 Soares tota
gb_est22:AI584636	+ 563.00	1117.03	4.2e-53	530	AI584636 fb81e05.y1 Zebrafish W
gb_est47:AM460650	- 556.00	1099.05	4.2e-52	731	AM460650 da29g11.x1 Xenla 13LiC
gb_est97:BG161868	- 522.00	1030.94	2.6e-48	687	BG161868 df69e06.x1 Xenopus lae
gb_est107:R21282	+ 521.50	1036.08	1.4e-48	408	R21282 Yg47d12.r1 Soares infant
gb_est51:AM767659	- 508.00	1002.87	9.7e-47	671	AM767659 da77a02.x1 Harland sta
gb_est40:AV668461	+ 501.00	992.36	3.7e-46	492	AV668461 AV668461 Sugano Kawa
gb_est40:AV668503	+ 499.00	988.82	5.9e-46	471	AV668503 AV668503 Sugano Kawa
gb_est71:BE188917	- 475.00	936.92	4.6e-43	623	BE188917 db61c05.x1 Wellcome C
gb_est43:AM197734	+ 472.50	927.27	1.6e-42	919	AM197734 da60g04.y1 Xenopus lae
gb_est43:AM128313	- 468.50	922.97	2.7e-42	666	AM128313 fc37g11.x1 Zebrafish W
gb_est21:AI545454	- 466.00	915.98	6.7e-42	784	AI545454 fb81e05.x1 Zebrafish W
gb_est49:AM640733	+ 463.50	918.43	4.9e-42	415	AM640733 b198d01.w1 Blackshear
gb_est101:BG487359	- 442.00	870.88	2.2e-39	583	BG487359 dad56d03.x1 Wellcome C
gb_est51:AM766054	+ 424.50	833.90	2.5e-37	664	AM766054 da81f06.y1 Harland sta
gb_est101:BG487338	- 421.00	828.25	5.1e-37	588	BG487338 dad56a04.x1 Wellcome C

gb_est75:BE506442 + 407.50 805.32 9.7e-36 405 | BE506442 db83b11.y1 Wellcome
gb_est101:BG486108 - 404.00 793.73 4.3e-35 593 | BG486108 dad60f10.x1 Wellcom
gb_gss5:CNS05PUV + 402.50 783.90 1.5e-34 1054 | AL348447 Tetraodon nigrovir
gb_est99:BG348623 + 395.50 782.18 1.9e-34 367 | BG348623 dad18c02.y1 Wellcom

seq_name: gb_est29:AL539668

seq_documentation_block:

LOCUS AL539668 974 bp mRNA EST 16-FEB-2001
DEFINITION AL539668 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF036YG04 5
prime, mRNA sequence.
ACCESSION AL539668
VERSION AL539668.1 GI:12869097
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 974)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..974
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF036YG04"
/clone_lib="LTI_FL013_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: PCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
PCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 223 a 269 c 305 g 177 t
ORIGIN

alignment_scores:

Quality: 1431.50 Length: 300
Ratio: 4.853 Gaps: 1
Percent Similarity: 98.333 Percent Identity: 96.667

alignment_block:

US-09-668-482-4 x AL539668 ..
Align seg 1/1 to: AL539668 from: 1 to: 974

54 pheglyglurthrleuglnmetvalleuglnarglyspheleuglnme 70
||||| ||| :::: ::||| ||||| |||||
77 TTTGGGTGAAC...AATGTCTGACCAAGCGAGGAGTTCCTGCAGAT 123
||||| ||| :::: ::||| ||||| |||||
70 tlysararglysthyrglypheietyrlysthrhisleupheglyargp 87
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124 GAAGCGCAGGAATATACGGCTTCATCTACAGACGATCTGTCGGCGGCC 173
87 rothrvalargvalmetglyalaaspsnvalargargileleuengly 103
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
174 CCACCGTACCGGTGATGGCGCGGACATGTGCGGCATCTTCCTCGGA 223
104 aspaspargleuvalservalhistrproalaservalargthillele 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

224 GA.GACCGGCTGGTGTGGTCCACTGGCCAGCGTGGTGGCACCATTCT 272
120 uGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgL 137
|||||
273 GGGATCTGGCTGCTCTCTAACTGCACGACTCTCGACACAAGCAGCGCA 322
137 ySLysValIleMetArgAlaPheSerArgLAlaLeuGluCysTyrVal 153
|||||
323 AGAAGTGATTAATGCGGGCCTTCAGCCGAGGA.CTCGAATGCTACGTG 371
154 ProValIleThrGluGluValGlySerSerLeuGluGlnTrpLeuSerCy 170
|||||
372 CCGGTGATCACCGAGGAAGTGGCAGACGCTGGAGCAGTGGCTGAGCTG 421
170 sGlyGluArgGlyLeuLeuValTyrProGluValLysArgLeuMetPhe 187
|||||
422 CGCGAGCGCGGCTCTGCTGTACCCCGAGGTGAAGCGCTCATGTCTCC 471
187 rGleAlaMetArgIleLeuLeuGlyCysGluProGlnLeuAlaGlyAsp 203
|||||
472 GAATCGCCATGGCATCTACTGGGCTGCGAACCCTGCGGGCGGAC 521
204 GlyAspSerGluGlnGlnLeuValGluAlaPheGluGluMetThrArgAs 220
|||||
522 GGGGACTCGAGCAGCAGCTTGTGAGGCGCTTCGAGGAATGACCCGCAA 571
220 nLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyM 237
|||||
572 TCTCTCTCGCTGCCCCATCGACGTGCCCTTCAGCGGGCTGTACCGGGCA 621
237 eTlyAlaArgAsnLeuIleHisAlaArgIleGluGlnAsnIleArgAla 253
|||||
622 TGAAGCGCGGGAACCTCATTCACGCCGCCATCGAGACAACATTCGCCGC 671
254 LysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCysLysAspAl 270
|||||
672 AAGATCTGCGGGCTGCGGGCATCCGAGCGGGCCAGGGCTGCAAGACGCG 721
270 aLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluArgLeuAspM 287
|||||
722 GCTGACGCTGTGATCGAGCAGCTCGTGGAGAGGGGAGAGCGGCTGACA 771
287 eTGlAlaLeuLysGlnSerSerThrGluLeuLeuPheGlyGlnHisGlu 303
|||||
772 TGCAAGCACATAAGCAATCTTCAACCGAATCTCTTTGGAGACACGAA 821
304 ThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrPr 320
|||||
822 ACCACGGCCAGTGCAACCATCTGTGATCACTTACCTGGGGCTTACCC 871
320 oHisValleuGlnLysValArgGluGluLeuLysSerLysGlyLeuLeu 337
|||||
872 ACATGTTCTCCAGAAAGTGCAGAGAAGCTGAAGAGTTACTTTT 921
337 ySLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeuGluGlnLeu 353
|||||
922 GCAAGAGCAATCAAGACAACAAGTTGACATGGAATTTTGAACACTTA 971

seq_name: gb_est29:AL532445

seq_documentation_block:

LOCUS AL532445 893 bp mRNA EST 13-FEB-2001
DEFINITION AL532445 LTI_NFL001_NBC4 Homo sapiens cdna clone CS0DM014YL13 5

prime, mRNA sequence.

ACCESSION AL532445

VERSION AL532445.1 GI:12795938

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 893)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES Location/Qualifiers

source

1..893

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DM014YL13"

/clone_lib="LTI_NFL001_NBC4"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 196 a 253 c 275 g 167 t 2 others
ORIGIN

alignment_scores:

Quality: 1376.00

Ratio: 4.745

Percent Similarity: 94.463

Percent Identity: 93.485

alignment_block:

US-09-668-482-4 x AL532445 ..

Align seg 1/1 to: AL532445 from: 1 to: 893

73 ArgLysTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVa 89
|||||
2 AGGAATACGGCTTCATCTACAAGACGATCTGTTCGGGGGCCACCCGCT 51
89 lArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAspA 106
|||||
52 ACGGTGATGGGCGCGGACAATGTGCGGCACTCTGCTCGGAGA.SACC 100
106 rGleuValSerValHisTrpProAlaSerValArgThrIleLeuGlySer 122
|||||
101 GGCTGCTGTGCTCCACTGGCCAGCGTGGTCCGACCATTTCTGGGATCT 150
123 GlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArg.LysLysV 139
|||||
151 GGCTGCCCTCTTAACCTGCACGACTCTCGCAACAAGCAGCGCAAGA.... 196
139 allIleMetArgAlaPheSerArgGluAlaLeuGluCysTyrValProVal 155
|||||
197AGCCGCGAGGCACTCGAATGCTACGTGCCGGTG 229
156 IleThrGluGluValGlySerSerLeuGluGlnTrpLeuSerCysGlyG 172
|||||
230 ATCACCGAGGAGAGTGGGCGAGCAGCGCTGAGCAGTGGCTGAGCTGGCG 279
172 uArgGlyLeuLeuValTyrProGluValLysArgLeuMetPheArgIleA 189
|||||
280 GCGCGGCTCTGCTCTACCCCGAGGTGAAGCGCTCATGTTCGGAATCG 329
189 lMetArgIleLeuLeuGlyCysGluProGlnLeuAlaGlyAspGlyAsp 205
|||||
330 CCATGCGCATCTACTGGGCTGCGAACCCTCACTGCGGGCGGCGAGCGGAC 379
206 SerGluGlnGlnLeuValGluAlaPheGluGluMetThrArgAsnLeuPh 222
|||||


```
380 TCCGAGCAGACGCTTGTGGAGCGCTTCGAGCAATGACCCGCAATCTCTT 429
222 eSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyMetLysA 239
|||||
430 CTCGCTGCCCATCGACGTGCCCTTCAGCGGCGCTGTACCGGGCATGAAG 479
239 IaArgAsnLeuIleHisAlaArgIleGluInAsnIleArgAlaLysIle 255
|||||
480 CGCGGAACCTCATCTCACGGCGCATCGACGACAACATTCGCCGCCAAGATC 529
256 CysGlyLeuArgAlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuG1 272
|||||
530 TGCGGGCTGCGGCATCCGAGGCGGCGCATGCAAGACCGCGCTGCA 579
272 nLeuLeuIleGluHisSerTrpGluArgGlyGluArgLeuAspMetGlnA 289
|||||
580 GCTGTTGATCGAGCACTCGTGGAGAGAGGAGCGGCTGGACATGCAGG 629
289 IaLeuLysGlnSerSerThrGluLeuLeu.PheGlyGlyHisGluThrTh 305
|||||
630 CACTAAAGCAATCTTCAACCGAACTCCTTTTGGAGACACAGAAACAC 679
305 rAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisV 322
|||||
680 GGCCAGTGCAGCCACATCTCTGATCATTACTGGGGCTTACCCACATG 729
322 aLeuGlnLysValArgGluGluLeuLysSerLysGlyLeuLeuCysLys 338
|||||
730 TTCTCCAGAAAGTGCAGAGAGAGCTGAAGAGTTAAGGTTTACTTTGCAAG 779
339 SerAsnGlnAspAsnLysLeuAspMetGluIleLeuGlnLeuLysTyr 355
|||||
780 AGCAATCAAGACAAAGTTGGACATGAAATTGTGAACA.CTTAAATA 828
355 rIleGlyCysValIleLysGluThrLeuArgLeuAsnProProValProG 372
|||||
829 CATCGGGTGTGTTAAGGAGAGACCTTCGACTGAATCCCAGT...TCAG 875
372 IyGlyPheArgValAla 377
|||||
876 GAGGGTTTCGGGTTGCT 892
```

seq_name: gb_est29:AL532444

seq_documentation_block:

LOCUS AL532444 1028 bp mRNA EST 13-FEB-2001
DEFINITION AL532444 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM014YL13 3
prime, mRNA sequence.

ACCESSION AL532444
VERSION AL532444.1 GI:12795937

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 1028)

AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES Location/Qualifiers

source 1..1028

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DM014YL13"

/clone_1lb="LTI_NFL001_NBC4"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(df) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com>

BASE COUNT 265 a 244 c 225 g 282 t 12 others

ORIGIN

alignment_scores: Quality: 1313.00 Length: 282
Ratio: 4.955 Gaps: 3
Percent Similarity: 93.972 Percent Identity: 91.489

alignment_block: US-09-668-482-4 x AL532444/rev ..

Align seg 1/1 to reverse of: AL532444 from: 1 to: 1028

```
217 MetThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLe 233
|||||
1005 ATGACCCGA..ATCTCTCTGCTGCCATCGACTGCCCT...TCAGCGGC 962
233 uTyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnA 250
|||||
961 TGTACCGGGCATGTAGCGCGCAACCTTAATTAAACGCGGCATCGAGCAGA 912
250 snIleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGly 266
|||||
911 AMATACGGGCCAAGATCTGCGGGCTGCGGCATCCGAGCGGGCCAGGCG 862
267 Cys.LysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyG 283
|||||
861 TGCAAAGACCGCGCTGCAGCTGTGATCGAGCACTCGTGGAGAGGGAG 812
283 IaArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPhe 299
|||||
811 AGCGGCTGACATGCAGGCACTAAAGCAATCTTCAACCGCACTCCTCTT 762
300 GlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLe 316
|||||
761 GGAGGACACGAAACACCGCCAGCTGCACCATCTCTGATMACTTACCT 712
316 uGlyLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerL 333
|||||
711 GGGGCTCTACCCACATGTTCTCCAGAAAGTGCAGAAAGAGCTGAAGAGTA 662
333 ysgLyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIle 349
|||||
661 AGGCTTACTTTGCCAAGAGCAATCAAGACAACAAGTTGGACATGGAAT 612
350 LeuGluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgIle 366
|||||
611 TTGAACAACCTTAATATACATCGGGTGTGTTAATTAAGAGAGACCTTCGACT 562
366 uAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGluL 383
|||||
561 GAATCCCCCAGTTCAGGAGGGTTTCGGTGTCTGAAGACTTTTGAAT 512
383 euAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCys 399
|||||
511 TAAATGATACAGATTCGCCAAGGGCTGGAATGTTATCTACAGTATCTGT 462
400 AspThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnPr 416
|||||
461 GATACTCATGATGTGGCAGAGATCTTCACCAACAAGGAAGAAATTAAATCC 412
416 oAspArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheI 433
|||||
```

411 TGACCGATTCATGCTGCTCACCACGAGAGATGCATCCAGGTTCAAGCCTCA 362
433 leProphEGlyGlyLeuArgSerCysValGlyLysGluPheAlaLys 449
361 TTCCATTTGGAGAGCCTTAGAGCTGTGTAGCAAGAATTGCAAAA 312
450 lLeLeuLysIlePheThrValGluLeuAlaArgHisCysAspTyrG1 466
311 ATTCTTCTCAAAATATTACAGTGAGCTGSCCAGCATGTGTACTGGCM 262
466 nLeuLeuAsnGlyProProThrMetLysThrSerProThrValTyrProV 483
261 GCTTCTAAATGAGACTTCCTACATGAAMCAGTCCACCGTGTATCCTG 212
483 aLAspAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
211 TGGACAATCTCCCTGCAAGATTCACCCCATTTCCATGGGGAATC 168
seq_name: gb_est48:AW513600

seq_documentation_block: 618 bp mRNA EST 03-MAR-2000
LOCUS AW513600
DEFINITION x047b06.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2707091 3'
similar to SW:CP26_HUMAN O43174 CYTOCHROME P450 26 ;, mRNA
sequence.
ACCESSION AW513600
VERSION AW513600.1 GI:7151678
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 618)
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco
High quality sequence stop: 440.

FEATURES
source
1. 618

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2707091"
/clone_lib="NCI_CGAP_Ut1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"

BASE COUNT 111 a 191 c 199 g 115 t 2 others
ORIGIN

alignment_scores:
Quality: 1045.00 Length: 206
Ratio: 5.148 Gaps: 0
Percent Similarity: 98.544 Percent Identity: 98.058

alignment_block:
US-09-668-482-4 x AW513600 ..
Align seg 1/1 to: AW513600 from: 1 to: 618

34 SerGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetG1 50
1 AGCGGCCGCGACCGCAGTTGTGCCCTCCCATGTGCCCGCGGACTATGG 50
50 yPheProPhEGlyGluThrLeuGlnMetValLeuGlnArgArgLysP 67
51 CTTCCCTCTTGGGGAACCTTGACAGATGGTACTGCAGCGAGAGAGT 100
67 heLeuGlnMetLysArgArgLysTyrGlyPheTleTyrLysThrHisLeu 83
101 TCCTGCAGATGAAGCGCAGGAATACGGCTTCACTACAAGACCATCTG 150
84 PheGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgT1 100
151 TTCGGGGCGGCCACCGTACGGGTGATGGCGCGGCAATGTGGCGGCAT 200
100 eLeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValA 117
201 CTTGCTCGAGAGACCGCGGCTGTGTGCTGCACCTGCCAGCGTCCGTGC 250
117 rGThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHis 133
251 GCACCATTTGGGATCTGGCTGCCCTCTTAACCTGCACGACTCCTCGCAC 300
134 LysGlnArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuG1 150
301 AAGCAGCGCAGAGAGGTGATTATGCGGGCTTCAGCGCGAGCACATCGA 350
150 uCysTyrValProValIleThrGluGluValGlySerSerLeuGluGlnT 167
351 ATGCTACGTGCGCGGTGATCACCGAGGAAGTGGCAGACGCTGGAGCAGT 400
167 rPLeuSerCysGlyGluArgGlyLeuValTyrProGluValLysArg 183
401 GGCTGAGCTGGCGGCGAGCGCGGCTCTGTCTTACCCCGAGGTGAAGCGC 450
184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLe 200
451 CTCATGTTCCGAATCGCCATGCGCATCTACTGTGCTGCGAAGCCCACT 500
200 uAlaGlyAspGlyAspSerGluGlnLeuValGluAlaPheGluGln 217
501 GCGGGCGGCGGAGCTCCGAGCAGCAGCTTGTGAGGCGCTTCGAGGAAA 550
217 eThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
551 TGACCCGCAATCTTCTCTGCTGCCCATCGACGTGCNCNTTCAGCGGCGCTG 600
234 TyrArgGlyMetLysAla 239
601 TACCGGGGCAATGAAGCN 618

seq_name: gb_est29:AL539667

seq_documentation_block: 847 bp mRNA EST 16-FEB-2001
LOCUS AL539667
DEFINITION AL539667 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF036YG04 3
prime, mRNA sequence.
ACCESSION AL539667
VERSION AL539667.1 GI:12869095
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. 847
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF036YG04"
/clone_lib="LTI_FL013_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: PCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 197 a 190 c 204 g 229 t 27 others
ORIGIN

alignment_scores:
Quality: 1024.00 Length: 231
Ratio: 4.995 Gaps: 0
Percent Similarity: 88.745 Percent Identity: 85.281

alignment_block:
US-09-668-482-4 x AL539667/rev ..

Align seg 1/1 to reverse of: AL539667 from: 1 to: 847

266 GLYCYSLYSASPALALEUGINLEUULEGIIHISSETRPGIUAARGGI 282
|||||
846 GGCTGCAAGACGCGCTGCAGCTKTGTATCGAGCACTCGTGAGAGGGG 797
282 YGIUARGLEUASPMETGIALALEULYSGLNSESERTHRGLULEUUP 299
|||||
796 AGAGCGGCTGACATGCGGCACTAAGCAATCTCAACCGAAGCTCTCT 747
299 HEGIYGLYHISGLUTHRTHRALASERIALAATHRSERLEUIETHTYR 315
|||||
746 TTGGAGGACACGAACACGCGCCAGTGCAGCCACATCTCTGATCATTAC 697
316 LEUGIYLEUYYRPROHISVAILEUGINLYSVALARGIUGIULEULYSSE 332
|||||
696 CTGGGGCTCCACCCACACGCTCCCGAGAAAGCGCGAAGAGCTGAAGAG 647
332 RLYSGIYLEULEUCYSLYSSERASNGINASPASNLYSLEUASPMETGLUI 349
|||||
646 TAAGGGTTTACTTTGCAGMGCAATCAAGACAACAGTGGACATGAAA 597
349 LELEUGIUGINLEULYSYRILEGLYCYSVALILEYSGIUTHRLEUARG 365
|||||
596 TTTTGGACAACACTCMAATMCCNCGGGTGTGTATTAAGGAGACCCCTCGA 547
366 LEUASNPROVALPROGLIYGLYPHEARGVALALEULYSTHRPHEGL 382
|||||
546 CTGAACCCCGCMGNCAGAGGGTTNCGGGNNGCTCCGAASMCCTTCTGA 497
382 ULEUASNGIYTYRGLNILEPROLYSGIYTRPASNVALILETYRSERIIEC 399
|||||
496 ATCAATGGATCCCMGANCCTCAAGGGCTGGAATGTTAICTACAGTATCT 447
399 YSASPTHRHISASPVAlAGIUILEPHETHRASNLYSGIUGIUPHEASN 415
|||||
446 GTGATACCATGATGTGGAGAGATCTTCACCAACAAGAGAAATTTAAC 397

416 PROASPARPHESERIALAPROHISPROGLUASPALASERARPHESERPH 432
|||||
396 CCTGACCGATTTCATGCTGCTTCACCCMGAGAGCGACCCMGCTCAGCCC 347
432 EIIEPROPHGLIYGLIYLEUARGSERCYSVALGLIYSGIUPHEALAL 449
|||||
346 CATCCATTGGAGAGGCGCTTAGGAGCTGTGTAGGCAAGAATTCGAMA 297
449 YSIIIEULEULYSILEPHERHVALGLULEUALARGHISCYSPTRP 465
|||||
296 AAATTCYCCCAAAATATTTCAGTGGAGCTGSCCAGCATTTGTGACTSG 247
466 GINLEUUAANGIYPROPROTHMETLYSTHRSEPROTHRVALTYRPR 482
|||||
246 CCSBRTCTAATGACCTCTACABGAAACAGNCCCGTGTATCC 197
482 OVALASPNLEUPROALAARGPHETHRHSIPHEHISGLIGLU 496
|||||
196 TGTGACCAATCTCCCTGCAAGATTCACCCATTCCTGGGAA 154

seq_name: gb_est72:BE236243

seq_documentation_block:
LOCUS BE236243 537 bp mRNA EST 10-JUL-2000
DEFINITION 143959 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE236243
VERSION BE236243.1 GI:9020961
KEYWORDS EST.
SOURCE
ORGANISM Sus scrofa
pig.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE 1 (bases 1 to 537)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 85 row: B column: 21
Seq primer: ATTTAGGTGACACTATAG.

FEATURES Location/Qualifiers

source 1. 537

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1PIG"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT 96 a 163 c 174 g 104 t
ORIGIN

alignment_scores:
Quality: 881.00 Length: 177
Ratio: 5.063 Gaps: 0
Percent Similarity: 98.305 Percent Identity: 93.785

alignment_block:
US-09-668-482-4 x BE236243 ..

Align seg 1/1 to: BE236243 from: 1 to: 537

```
25 IleLysLeuTrpAspLeuTyrCysValSerGlyArgAspArgSerCysAl 41
|||||
1 ATCAAACTCTGGACCTGTACTGCGTGAGACGCGGAGCCGACGCTGCAC 50
41 aleuProLeuProGlyThrMetGlyPheProPheGlyLutHrL 58
|||||
51 CCTTCCTTGGCCCCCTGGAACATATGGGCTTCCCTCTTGGGAGACAT 100
58 euglMetValLeuGlnArgArgLysPheLeuGlnMetLysArgArgLys 74
|||||
101 TGCAGATGCTGCTACAGCGAAGAGTCTCTGCAGATGAAGCGCAGGAAA 150
75 TyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrValArgVa 91
|||||
151 TACGGTTTCATCTACAGACGATCTGTCTGGGAGGCCACGGTGGGGT 200
91 lmetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAspArgLeuV 108
|||||
201 GATGGGTGCAGACACAGCTGCGGGCGCATCTGTCTCGGGGGAACACCGGCTCG 250
108 alSerValHisTrpProAlaSerValArgThrIleLeuGlySerGlyCys 124
|||||
251 TGTCTGCTCAGCTGCGCGGCTGCGGTGCGCAGATCTGGGCTGTGGCTGC 300
125 LeuSerAsnLeuHisAspSerSerHisLysGlnArgLysValIleLeu 141
|||||
301 CTCTCCAACCTGCAGACTCTCTGCACAAAGCAGCGCAAGAGTGTAT 350
141 tArgAlaPheSerArgGluAlaLeuGluCysTyrValProValIleThrG 158
|||||
351 GCAGGCTTCAAGCCGAGGCGCTCCAGTGTCTGCTGCGGTGATCGCAG 400
158 lgluValGlySerSerLeuGluGlnTrpLeuSerCysGlyLutArgGly 174
|||||
401 AGGAAGTGGACAGTGTGCTGAGACAGTGGCTGAGCTGCGGAGAGCGGCG 450
175 leuLeuValTyrProGluValLysArgLeuMetPheArgIleAlaMetAr 191
|||||
451 CTCTGCTTACCCCCAGGTGAACGCTCATGTTCGCCATCGCCATGCG 500
191 gileLeuLeuGlyCysGluProGlnLeuAla 201
|||||
501 CATCCTGCTGGGCTGCGAGCCCGGCTGGCG 531
```

seq_name: gb_est87:BF385034

seq_documentation_block: 539 bp mRNA EST 27-NOV-2000
LOCUS BF385034 602045632F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4195422 5',
DEFINITION mRNA sequence.

ACCESSION BF385034
VERSION BF385034.1 GI:11366326
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 539)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM9530 row: 0 column: 07
High quality sequence stop: 539.

FEATURES
source location/Qualifiers
1..539

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4195422"
/clone_lib="NCI_CGAP_L19"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 86 a 160 c 181 g 112 t
ORIGIN

alignment_scores:
Quality: 852.00 Length: 171
Ratio: 5.041 Gaps: 0
Percent Similarity: 98.830 Percent Identity: 94.152

alignment_block:
US-09-668-482-4 x BF385034 ..

Align seg 1/1 to: BF385034 from: 1 to: 537

```
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
25 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCGCTGCACTTGCAGCTTGTGCTGCC 74
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
75 GCTGCTGCTTCTCTGCGCGCGCTCAAGCTGTGGACCTGTACTGTGTGA 124
34 ergLysArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
125 GCAGCCGCGATCGCAGCTGCGCCCTCCCTTGGCCCGCGGTACCATGGGC 174
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
175 TTCCCATCTTCTTGGGGAACATTCAGATGTGCTTCAAGCGGAGAGAGTT 224
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
225 TCTGCAGATGAAGCGCAGGAATAACGGCTTCATCTACAAGACGCATCTGT 274
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
275 TTGGGCGGCCCCACGGGTGCGGTGATGGGCGCGGATATGTGCGGCGCATC 324
101 leuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
325 TTGCTGGAGAGACACCGGTTGTGTGCTGCTGACCTGGCCCGCGTGGTGCG 374
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
375 CACCATCTGCGCGCTGCTGCTCTCCAACTGCACAGATTCTCCGCACACA 424
134 ysgLysArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
425 AGCAGCGAAGAGAGTGTATGACAGCCCTTCAGCCGCGAGGACTCCAG 474
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
|||||
475 TGCTACGTGCGCCGTGATCGTGAAGAGATCAGCAGTTGTCTGGAGCAGTg 524
167 pleuSerCysGly 171
|||||
525 GCTAAGCTGCGGC 537
```



```

seq_name: gb_est89:BF533769

seq_documentation_block:
LOCUS      BF533769          564 bp      mRNA          EST          11-DEC-2000
DEFINITION 602073982F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4210893 5',
            mRNA sequence.
ACCESSION   BF533769
VERSION     BF533769.1  GI:11621132
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 564)
REFERENCE   1  (bases 1 to 564)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM9779 row: c column: 22
            High quality sequence stop: 564.

FEATURES
    source
        1..564
            location/Qualifiers
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:4210893"
                /clone_lib="NCI_CGAP_Li9"
                /lab_host="DH10B (T1 phage-resistant)"
                /note="Organ: liver; Vector: PCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.9 kb. Constructed by Life
                Technologies. Note: this is a NCI_CGAP library."

BASE COUNT      89 a      172 c      189 g      114 t

alignment_scores:
    Quality:      852.00      Length:      171
    Ratio:        5.041      Gaps:      0
    Percent Similarity: 98.830      Percent Identity: 94.152

alignment_block:
US-09-668-482-4 x BF533769 ..

Align seg 1/1 to: BF533769 from: 1 to: 564

```

```

250  TCTGCAGATGAAGCGCAGAGAAATACGGCTTCATCTACCAAGACGATCTGT 299
84   heGIyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
300  TTGGGGCGGCCACCGGTGGCGGTGATGGCGCGGATATGTGGCGGCATC 349
101  LeuLeuGIyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
350  TTGCTGGGAGAGACACCGGTGGTGTCTGGGTGCACCTGGCCCGGTGGCG 399
117  gThrIleLeuGIySerGIySerGIySerLeuSerAsnLeuHisAspSerHisL 134
400  CACCATCTGGCGCGCTGGCTGCTCTCCAACTGCACGATTCTTCGCACA 449
134  ysGIyArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
450  AGCAGCGAAAGAAAGGTGATTATGCAGGCCCTTCAGCCGCGACACTCCAG 499
151  CysTyValProValIleThrGluGluValGIySerSerLeuGluInTr 167
500  TGCTACGTGCCCGCGTGCATCGTCGAGAACTCAGCAGTGTGTGAGCAGTG 549
167  pleuSerCysGly 171
550  GCTAAGCTGGCGC 562

seq_name: gb_est71:BE189825

seq_documentation_block:
LOCUS      BE189825          696 bp      mRNA          EST          07-AUG-2000
DEFINITION db61c05.y1 Wellcome CRC psk egg Xenopus laevis cDNA clone
            IMAGE:3377480 5' similar to gb:gb|AF057566.1|AF057566 Xenopus
            laevis retinoic acid converting enzyme (XENOPUS);, mRNA sequence.
            BE189825
            BE189825.2  GI:9729548
ACCESSION   BE189825
VERSION     BE189825.2
KEYWORDS    EST.
SOURCE      African clawed frog.
ORGANISM    Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
            Xenopodinae; Xenopus.
            1 (bases 1 to 696)
REFERENCE   1  (bases 1 to 696)
AUTHORS     Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
            Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Rilter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
            Washu Xenopus EST project, 1999
            Unpublished (1999)
            On Jun 22, 2000 this sequence version replaced gi:8668718.
            Other_ESTs: db61c05.x1
            Contact: Sandy Clifton, Ph.D.
            Washu Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B.
            Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
            University Genome Sequencing Center
            Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 483.

FEATURES
    source
        1..696
            location/Qualifiers
                /organism="Xenopus laevis"
                /db_xref="taxon:8355"
                /clone="IMAGE:3377480"
                /clone_lib="Wellcome CRC psk egg"
                /tissue_type="egg"
                /lab_host="DH10B (phage-resistant)"
                /note="Vector: pBluescript SK-; Site_1: NotI; Site_2:

```

ECORI; cDNAs were oligo-dT primed and directionally cloned. Library was constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute).
Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT 156 a 200 c 202 g 138 t

ORIGIN

alignment_scores: Quality: 840.00 Length: 235
Ratio: 4.118 Gaps: 4
Percent Similarity: 86.809 Percent Identity: 71.064

alignment_block:
US-09-668-482-4 x BE189825 ..

Align seg 1/1 to: BE189825 from: 1 to: 696

43 PROLEUPROPROGLYTHRMETGLYPHEPROPHHEGLYGLUTHRLEUG1 59
|||||
8 CCACCTGCCCGCCGCACTATGGGGCTGCCCTTCTCGGGAGACTCTGCA 57
59 nmetValLeuGlnArgArgLysPheLeuGlnMetLysArgArgLysTyrG 76
|||||
58 AATGGTGTGAGAGAGCGCAAGTCTCCAAATGAAGCGTAGAAAGTAGC 107
76 lypheileTyrLysThrHisLeuPheGlyArgProThrValArgValMet 92
|||
108 GTGCGATCTACAGACGCGATCTGTGCGTAGCCCCACGCGCGCTCACG 157
93 GLYALAAspAsnValArgArgLysLeuLeuGlnLysAspAspArgLeuValSe 109
|||||
158 GGGCGAGAGAACGTTCGCCAGATCCTATTGGGGAGACACAAAGCTGTGTC 207
109 rValHisTrpProAlaSerValArgThrIleLeuGlySerGlyCysLeuS 126
|||||
208 GGTGCACTGGCGCGGCTCGGTGCGCACGATCCTAGGGCGGCTGTCTGT 257
126 eRAsnLeuHisAspSerSerHisLysGlnArgLysLysValIleMetArg 142
|||||
258 CCAACCTGCACGACTCTGAGCACAACTACCAAGAAAGTAGATTGCACAA 307
143 AlaPheSerArgGluAlaLeuGluCysTyrValProValIleThrGluG1 159
|||||
308 GCCTTCTCCGAGATGCCCTGCAATTAACGTGCCGAGATGGAAGAAGA 357
159 uValGlySerSerLeuGlnTrpLeuSerCysGlyLysArgGlyLeuL 176
|||||
358 GGTGAGACGCTCTGTAACCTGTGGCTGCAGAGCGGCCCC...TGCCTGC 404
176 euValTyrProGluValLysArgLeuMetPheArgIleAlaMetArgIle 192
|||||
405 TGATGTATCCCGCATCAAGCGACTGATGTTCGCAATGCCATGAAGCTC 454
193 leuLeuGlyCysGluProGlnLeuAlaGlyAspGlyAspSerGluGlnG1 209
|||||
455 CTGCTCGGCTGCGATCCCGAGCGCATG...GACAGCCAACATGAGGAGAC 501
209 nLeuValGluAlaPheGluGlnMetThrArgAsnLeuPheSerLeuProI 226
|||||
502 GCTGCTCGAAGCCTTCGAGGAATGACCCGAAATCTTCTCTGTGCTTA 551
226 leaspValProPheSerGlyLeuTyrArgGlyMetLysAlaArgAsnLeu 242
|||||
552 TTGATGTGCCATGTAGCGGCTCTACCGGGGTCTGCGGCTAGAGACCTT 601
243 lLeHisAlaArgIleGluGlnAsnIleArgAlaLysIleCysGlyLeuAr 259
|||||
602 ATTCATGCGCGAATTGATGAACATCGAAGAGAAAGCTG.....CTAAG 645
259 gAlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeuLeuIleG 276
|||
646 G.....GAACACAGACGATTAATTGCCGGGATGCCCTGCAGCTGTGATTG 689

276 luhis 277
:::
690 ACTAT 694

seq_name: gb_est4:AA239785

seq_documentation_block:

LOCUS AA239785 474 bp mRNA EST 03-MAR-1997
DEFINITION mx80a03.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:692620 5'
similar to TR:G1001252 G1001252 HYPOTHETICAL 50.6 KD PROTEIN ;,
mRNA sequence.

ACCESSION AA239785
VERSION AA239785.1 GI:1863825

KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 474)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HM1 Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HM1 Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:426180

Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 461.
Location/Qualifiers

FEATURES
source

1. 474
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:692620"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGGAGCGCGCGCAATCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT 135 a 99 c 118 g 122 t

ORIGIN

alignment_scores: Quality: 823.00 Length: 158
Ratio: 5.310 Gaps: 0
Percent Similarity: 98.101 Percent Identity: 96.203

alignment_block:
US-09-668-482-4 x AA239785 ..

Align seg 1/1 to: AA239785 from: 1 to: 474

324 GlnLysValArgGluLeuLysSerLysGlyLeuLeuCysLysSerAs 340
|||||
1 CAGAAAGTTCGAGAAGAGATTAAGAGCAAGGCTTACTTTCACAGAGCAA 50

```
340 nglnaspasnlysluaspmetglutlleugluInleuIysTyrIleg 357
|||||
51 TCAAGACACAAAGTTAGACATGGAACCTTGGAAACAGCTTAATAATACACTG 100
|||||
357 lYcysValIleIySGluThrIeuArgIeuAsnProProValProGIyGIy 373
|||||
101 GGtGTGTCAATTAAGAGAGACCCTGGATGATCCTCCGGTTCAGAGAGG 150
|||||
374 pheArgValAlaIeulysThrPheglIuIeuAsnglyTyrGlnIleProly 390
|||||
151 TTTCGGGTGTCTCTGAAGACTTTTGAGCTGAATGGATGATACCAGATCCCAA 200
|||||
390 sGIYTrpasnValIleTyrSerIleCysAspThrHisAspValAlaGluI 407
|||||
201 GGGCTGAATGTTATTACAGTATCTGTGACACACCACGATGTGGCAGATA 250
|||||
407 IepheThrasnlySGluGluPheasnProasparGpPheSerAlaProHis 423
|||||
251 TCTTCACTAACAAAGAGGAATTTAATCCGACCGCTTATAGTGCTCAT 300
|||||
424 ProgluAspAlaSerArgPheSerPheIleProPheglYglYglYleuAr 440
|||||
301 CCAGAGGATGCTTCCCGGTTCAGCTTCATTCATTTGGAGAGGCGCTTCG 350
|||||
440 gSerCysValglYlySGluPheAlaIalysIleuIeuIysIlePheThry 457
|||||
351 GAGCTGTGTAGGCAAGAGATTGCAAAAATCTCTTAAGATATTACAG 400
|||||
457 aIGluIeuAlaArgHisCysAspTrpGlnIeuIeuAsnglyProProThr 473
|||||
401 TGGAGCTGGCTAGGCACTGTGATTGGCAGCTTCTAAATGGACCTCCTACA 450
|||||
474 MetLysThrSerProThrValTyr 481
|||||
451 ATGAAGACAAAGCCCACTGTGTAC 474
|||||
```

seq_name: gb_est85:BF236872

seq_documentation_block:

LOCUS BF236872 545 bp mRNA EST 14-NOV-2000
DEFINITION 602027980F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4163202 5',
mRNA sequence.

ACCESSION

BF236872

VERSION

BF236872.1 GI:11150789

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1 (bases 1 to 545)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM9446 row: p column: 19
High quality sequence start: 2
High quality sequence stop: 545.
Location/Qualifiers
1..545

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4163202"
/clone_lib="NCI_CGAP_Li9"

/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 87 a 165 c 184 g 109 t
ORIGIN

alignment_scores:
Quality: 769.00 Length: 173
Ratio: 4.633 Gaps: 4
Percent Similarity: 95.954 Percent Identity: 89.017

alignment_block:
US-09-668-482-4 x BF236872 ..

Align seg 1/1 to: BF236872 from: 1 to: 545

```
1 MetGIYleuProAlaIeulEuAlaSerAlaIeuCysThrPheValIeuPr 17
|||||
35 ATGGGGCTCCCGCGGCTGTGGCCAGTGGCGCTGTGCACCTTGTGTGCC 84
|||||
17 OleuIeuIeuPheIeuAlaIalIySleuTrpasIeuTyrCysValS 34
|||||
85 GGTGCTGCTCTTCTGGCGGCGCTCAAGCTCTGGGACCTGTACTGTGTGA 134
|||||
34 erGIYArgAsparGserCysAlaIeuProIeuProProGIYThrMetGIy 50
|||||
135 GCAGCCGGCATCGCAGCTGGCCCTCCCTTGCCCCCGGTACCATGGGC 184
|||||
51 PheProPhePheglYgluThrIeuIeuMetValIeuGlnArgArgLysp 67
|||||
185 TTCCCATTTCTTGGGGA...AACATGCAGATGGTGTCTTACAGCGAGGAAGT 231
|||||
67 heIeuGlnMetLysArgArgLysTyrGlyPheIleTyrIlyThrHisIeu 83
|||||
232 TTCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAAAGCATCTCG 281
|||||
84 PheglY.ArgProThrValArgValMetGIYAlaAspAsnValArgArgI 100
|||||
282 TTGGGCGGGCC...ACGGTGGCGGTGATGGCGCGGATATGTGCGCGCA 328
|||||
100 IeIeuIeuGIYAspAsparGleuValSerValHisTrpProAlaSerVal 116
|||||
329 TCTTGCTGGGAGACACACCGGTGTGTCGTGCACATGGCCCGCGGTG 378
|||||
117 ArgThrIleIeuGlySerGIYcysIeuSerAsnIeuHisAspSerSerHi 133
|||||
379 CGCACCATCTGTGGCGCTGTGCTCTTCAACCTGCACGATTCCTCGCA 428
|||||
133 sLYSGlnArgLysValIleMetArgAlaPheSerArgGluAlaIeuG 150
|||||
429 CAAGCAGCGAAAGAGGTGATTATGACAGCGCTTCAAGCGGAGGCACTCC 478
|||||
150 lUcysTyrValProValIleThrGluGluValGlySerSerIeuGluGln 166
|||||
479 AGTGCTACGTGCCGTGATGCTGAGGAAGTCAGCAGTTGTGTGACGACG 528
|||||
167 TrpIeuSerCysGIy 171
|||||
529 TGGCTAAGCTGCGGC 543
|||||
```

seq_name: gb_est82:BF055367

seq_documentation_block:

LOCUS BF055367 646 bp mRNA EST 16-OCT-2000
DEFINITION 7378f12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:3392591 3' similar to SW:CP26_HUMAN 043174 CYTOCHROME P450 26

ACCESSION

BF055367

VERSION

BF055367.1 GI:10809263

KEYWORDS

EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 646)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 450.
FEATURES
source
1. 646
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3392591"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and cloneids: Soares NbHSF pool 1:
309384-310919, 323208-325895 Soares Nb2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares Nb2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NbHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NbHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 202 a 129 c 120 g 194 t 1 others
ORIGIN
alignment_scores:
Quality: 763.00 Length: 149
Ratio: 5.336 Gaps: 0
Percent Similarity: 95.973 Percent Identity: 93.960
alignment_block:
US-09-668-482-4 x BF055367/rev ..
Align seg 1/1 to reverse of: BF055367 from: 1 to: 646
349 ltleuGlInleuLysTyrIleGlyCysValIleLysGluThrLeuAr 365
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
646 ATTTGGCCACCTTAATACCTCGGGTGTGTATTAAGAGACCTTNG 597
365 gLeuAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheG 382
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
596 ACTGAATCCCCAGTTCACGAGAGGGTTTCGGGTGCTTGAAGACTTTTA 547
382 luleuAsnGlyTyrGlnIleProLysGlyTTPasnValIleTyrSerIle 398
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||
546 ATTAATAATGATACACAGATTCCTCAAGGGCTGAATGTATCTACAGTATC 497
399 CysAspThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAs 415
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
496 TGTGATACTCATGATGTGGAGAGATCTTCACCAACAGAGAATTAA 447
415 nProAspArgPheSerAlaProHisProGluAspAlaSerArgPheSerP 432
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
446 TCCTGACCGATTCACTGCTCCACCCAGAGAGATGATCCAGGTTCAAGCT 397
432 heileProPheGlyGlyLeuArgSerCysValGlyLysGluPheAla 448

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
396 TCATTCATTTGGAGAGAGCCCTTAGAGAGCTGTGTAGGCAAGAATTGCA 347
449 lvsIleleuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTr 465
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
346 AAAATTCTTCTCAAAATATATTACAGTGGAGCTGGCCAGGCAATGTGACTG 297
465 pGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrValTyrP 482
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
296 GCAGCTCTTAATGAGACCTCTACATGAAACACGATCCACCGTGTATC 247
482 roValAspAsnLeuProAlaArghPheThrHisPheHisGlyGluIle 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 CTGTGGACAATCTCCCTGCAAGATTCACCCATTTCCATGGGGAATC 200
seq_name: gb_est43:AW174347
seq_documentation_block:
LOCUS AW174347 720 bp mRNA EST 16-NOV-1999
DEFINITION f142a06.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
2640274 5' similar to SW:CP26_BRARE P79739 CYTOCHROME P450 26 ;,
mRNA sequence.
ACCESSION AW174347
VERSION AW174347.1 GI:6440295
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter
,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
Washu Zebrafish EST Project 1999
Unpublished (1999)
TITLE Other_ESTs: f142a06.x1
JOURNAL
COMMENT Contact: S.L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: T3 ET from Amersham
High quality sequence stop: 478.
FEATURES
source
1. 720
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="2640274"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACGTGTCG);
Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TGTGGCCCTACTG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTCG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTCTGCTCTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGAGACA."
BASE COUNT 172 a 176 c 196 g 176 t

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
||| ||| ::|||:::|||||:::|||||:::|||||
35 ATGATCTGTATACTCTGCTACTAGCGCTGTGCACCTTGGCGTCCC 84
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrcysValS 34
|:::|||||:::|||||:::|||||:::|||||:::|||||
85 AGTGCTACTCTCTGCTCACCGCTGCCAAACCTGTGGAATGTACTGTCTGA 134
34 ergLYarGAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
:::||||| :::||||| |||||:::|||||:::|||||
135 GGCGCAAGATCCACCTCGCAACCCACTGCCCCCGGCACCTATGGGG 184
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
:::|||||:::|||||:::|||||:::|||||:::|||||
185 CTGCCCTCTCTCGGGGAGAGACTCTGCAATGTGCTGCAGAGCGCAAGTT 234
67 eleuGlnMetLysArgArgLysTyrglyPheIleTyrlYsthrHisLeuP 84
|||||:::|||||:::|||||:::|||||:::|||||
235 CCTCCAATGAAGCGGTAGAAAGTACGGTGCATCTACAAGACGCATCTGT 284
84 heGLyargProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
||||| |||||:::|||||:::|||||:::|||||
285 TCGGTAGCCCTACGGTGCCTCAGCGGCGCAGAACGTTCCGCGAGATC 334
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||:::|||||:::|||||:::|||||:::|||||
335 CTATTGGGGGAGACACAAAGCTGTGCTGCTGCTGCGCGCGCTGGTGGCG 384
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||:::|||||:::|||||:::|||||:::|||||
385 CACGATCTAGGGGCGCGGTGTCTGTCCAACTGCACGACTCTGAGCACA 434
134 ysglnArgLysLysValIleMetArgAlaPheSerArgGlnAlaLeuGlu 150
|| |||||:::|||||:::|||||:::|||||:::|||||
435 AGTACACCAAGAAAGTATGCACAAAGCTTCTCCGAGATGCCCTCGAC 484
151 CysTyValProValIleThrGluGluValGlySerSerLeuGlnGlnTr 167
|||||:::|||||:::|||||:::|||||:::|||||
485 AATTATGTGCCGAGATGAGAGAGAGGTGAGACGCTGTGTAACCTGTG 534
167 pleuSerCysGlyGluArgGlyLeuLeuValTyPrGluValLysArgL 184
|||||:::|||||:::|||||:::|||||:::|||||
535 GCTGCAGAGCGGCCCC...TGCGTGTGATGTATCCCGCATCAAGCGAC 581
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGln 199
|||||:::|||||:::|||||:::|||||:::|||||
582 TGATGATTGCCATTTGCATGAGGCTGCTCCGCTGCATCCCCCAG 628

seq_name: gb_est23:AI667038

seq_documentation_block: mRNA EST 18-MAY-1999
LOCUS AI667038 640 bp
DEFINITION fc24h03.y1 Zebrafish Washu MPIMG EST Danio rerio cDNA 5' similar to
SW:CP26_BRARE P79739 CYTOCHROME P450 26 ;, mRNA sequence.
ACCESSION AI667038
VERSION AI667038.1 GI:4805394
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio

REFERENCE
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.

TITLE Washu Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Ressourcenzentrum/Primardatenbank, Berlin, Germany (web address:
www.rzpd.de)
Possible reversed clone: similarity on wrong strand
Seq primer: T3 ET from Amersham
High quality sequence stop: 474.

FEATURES

source
1. 640
Location/Qualifiers

/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish Washu MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="Xl1-blue MRF"

/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dt)15 primer
[5'pGACTAGTTCTAGATCGCAGCGGCGCCCTTTTCTTTTCTT3'];
double-stranded cDNA was ligated into the Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). CDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

BASE COUNT 156 a 162 c 167 g 155 t
ORIGIN

alignment_scores:

Quality: 691.50 Length: 197
Ratio: 4.044 Gaps: 2
Percent Similarity: 86.802 Percent Identity: 68.020

alignment_block:
us-09-668-482-4 x AI667038 ..

Align seg 1/1 to: AI667038 from: 1 to: 640

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||:::|||||:::|||||:::|||||
62 ATGGGGCTGTACACCTTATGTGTCACCTTCTCTGCACCATCGTGTACC 111
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrcysValS 34
|:::|||||:::|||||:::|||||:::|||||
112 CGTTTACTCTTCTCGCGCGGTGAAGTTGTGGAGATGTTAATGATCC 161
34 ergLYarGAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
||| :::||||| |||||:::|||||:::|||||
162 GACGAGTCGATCCGAACTGCAAGATCCTTACCGCAGTACCATGGGC 211
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
:::|||||:::|||||:::|||||:::|||||
212 TTGCCGTTCAATGGAGAAAGCTCCAGCTGATCTCCACAGAAAGAAAGTT 261
67 eleuGlnMetLysArgArgLysTyrglyPheIleTyrlYsthrHisLeuP 84

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:27:05 ; Search time 88.14 Seconds
(without alignments)
126.891 Million cell updates/sec

Title: US-09-668-482-32
Perfect score: 2612
Sequence: 1 MGLPALLASALCTFVLPLLL.....PTVYVVDNLPARFTYFGDI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2612	100.0	497	4 US-08-882-164D-32	Sequence 32, Appl
2	2437	93.3	497	3 US-08-724-466B-4	Sequence 4, Appli
3	2437	93.3	497	4 US-08-882-164D-4	Sequence 4, Appli
4	1734	66.4	492	3 US-08-724-466B-2	Sequence 2, Appli
5	1734	66.4	492	4 US-08-882-164D-2	Sequence 2, Appli
6	450	17.2	472	2 US-08-622-166A-2	Sequence 2, Appli
7	450	17.2	472	2 US-08-622-166A-4	Sequence 4, Appli
8	358.5	13.7	511	4 US-08-991-677-4	Sequence 2, Appli
9	337	12.9	520	2 US-09-091-432-2	Sequence 16, Appl
10	300	11.5	576	3 US-08-948-564-16	Sequence 50, Appl
11	284.5	10.9	510	3 US-08-948-564-4	Sequence 4, Appli
12	270.5	10.4	496	1 US-08-313-075A-50	Sequence 68, Appl
13	262.5	10.0	500	4 US-09-292-768-68	Sequence 4, Appli
14	261.5	10.0	500	4 US-09-292-768-4	Sequence 70, Appl
15	261.5	10.0	500	4 US-09-292-768-70	Sequence 30, Appl
16	258.5	9.9	476	1 US-08-313-075A-30	Sequence 66, Appl
17	256.5	9.8	500	3 US-08-881-784-9	Sequence 2, Appli
18	255.5	9.8	513	3 US-08-948-564-2	Sequence 6, Appli
19	255.5	9.8	513	3 US-08-948-564-6	Sequence 6, Appli
20	249	9.5	510	3 US-08-606-505B-66	Sequence 2, Appli
21	249	9.5	510	4 US-09-616-990-66	Sequence 2, Appli
22	242	9.3	500	2 US-08-314-601-2	Sequence 2, Appli
23	242	9.3	500	5 PCT-US95-13051-2	Sequence 25, Appl
24	240	9.2	504	1 US-08-457-274A-25	Sequence 25, Appl
25	240	9.2	504	5 PCT-US95-05758-25	Sequence 64, Appl
26	239	9.1	496	4 US-09-292-768-64	Sequence 18, Appl
27	238.5	9.1	509	3 US-08-948-564-18	

28	237.5	9.1	490	1 US-08-201-118-3	Sequence 3, Appli
29	237.5	9.1	490	1 US-08-201-118-9	Sequence 9, Appli
30	237.5	9.1	490	2 US-08-238-821B-3	Sequence 3, Appli
31	237.5	9.1	490	2 US-08-238-821B-9	Sequence 9, Appli
32	237.5	9.1	490	5 PCT-US95-05744-3	Sequence 9, Appli
33	237.5	9.1	490	5 PCT-US95-05744-9	Sequence 10, Appl
34	237.5	9.1	532	3 US-08-948-564-10	Sequence 1, Appli
35	235	9.0	496	3 US-08-881-784-1	Sequence 2, Appli
36	235	9.0	496	4 US-09-292-768-2	Sequence 66, Appl
37	235	9.0	496	4 US-09-292-768-66	Sequence 6, Appli
38	234.5	9.0	496	4 US-09-172-339-6	Sequence 1, Appli
39	231.5	8.9	490	1 US-08-201-118-1	Sequence 1, Appli
40	231.5	8.9	490	2 US-08-238-821B-1	Sequence 1, Appli
41	231.5	8.9	490	5 PCT-US95-05744-1	Sequence 13, Appl
42	231	8.8	490	1 US-08-201-118-13	Sequence 13, Appl
43	231	8.8	490	2 US-08-238-821B-13	Sequence 13, Appl
44	231	8.8	490	5 PCT-US95-05744-13	Sequence 2, Appli
45	228.5	8.7	517	1 US-08-457-274A-2	

ALIGNMENTS

RESULT 1
US-08-882-164D-32
Sequence 32, Application US/08882164D
Patent No. 6306624
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-32

Query Match 100.0%; Score 2612; DB 4; Length 497;
Best Local Similarity 100.0%; Pred. No. 5.5e-276;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPLLLFLAALKLMDLYCVSSRDSCALPLPPGTMGPFPGFTLQM 60

```
|||||
Db 1 MGLPALLASALCTFVLPULLFLAALKLMDLYCVSSRDRSCALPLPPGTMGFFFGETLQM 60
QY 61 VLQRRKFLQMKRRKGYFYKTHLFGPPTVRVMGADNVRRIILGHEHRLVSVHWPASVRTL 120
Db 61 VLQRRKFLQMKRRKGYFYKTHLFGPPTVRVMGADNVRRIILGHEHRLVSVHWPASVRTL 120
QY 121 GAGCLSNLHDSHKQRRKVVIMQAFSREALQCYLVIAEEVSSCLEQWLSCGEGLLVYPE 180
Db 121 GAGCLSNLHDSHKQRRKVVIMQAFSREALQCYLVIAEEVSSCLEQWLSCGEGLLVYPE 180
QY 181 VKRLMFRIAMRILIGCEPPAGGEGDEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
Db 181 VKRLMFRIAMRILIGCEPPAGGEGDEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
QY 241 NLIHARIEENIRAKIRRLQATEPDGCKDALQLLIEHSEMERGERLDMQALKQSTELLFG 300
Db 241 NLIHARIEENIRAKIRRLQATEPDGCKDALQLLIEHSEMERGERLDMQALKQSTELLFG 300
QY 301 GHETTASATSLITYLGLYPHVLQKVREETIKSKGLCKSNQDNKLDMEITLQLYIGCVI 360
Db 301 GHETTASATSLITYLGLYPHVLQKVREETIKSKGLCKSNQDNKLDMEITLQLYIGCVI 360
QY 361 KETLRNLPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADIFTNKEEFPDRFI 420
Db 361 KETLRNLPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADIFTNKEEFPDRFI 420
QY 421 VPHPEDASRFSFIPEGGLRSCVKGFEAKILLKFTVELARHCDMQLNGPPTMTSPTV 480
Db 421 VPHPEDASRFSFIPEGGLRSCVKGFEAKILLKFTVELARHCDMQLNGPPTMTSPTV 480
QY 481 YPVDNLPARFTYFOGDI 497
Db 481 YPVDNLPARFTYFOGDI 497
```

```
RESULT 2
US-08-724-466B-4
; Sequence 4, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-724-466B-4

Query Match          93.3%; Score 2437; DB 3; Length 497;
Best Local Similarity 93.4%; Pred. No. 6,7e-257;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPULLFLAALKLMDLYCVSSRDRSCALPLPPGTMGFFFGETLQM 60
Db 1 MGLPALLASALCTFVLPULLFLAALKLMDLYCVSGDRSCALPLPPGTMGFFFGETLQM 60
QY 61 VLQRRKFLQMKRRKGYFYKTHLFGPPTVRVMGADNVRRIILGHEHRLVSVHWPASVRTL 120
Db 61 VLQRRKFLQMKRRKGYFYKTHLFGPPTVRVMGADNVRRIILGDDRLVSVHWPASVRTL 120
QY 121 GAGCLSNLHDSHKQRRKVVIMQAFSREALQCYLVIAEEVSSCLEQWLSCGEGLLVYPE 180
Db 121 GSGCLSNLHDSHKQRRKVVIMQAFSREALQCYLVIAEEVSSCLEQWLSCGEGLLVYPE 180
QY 181 VKRLMFRIAMRILIGCEPPAGGEGDEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
Db 181 VKRLMFRIAMRILIGCEPPAGGEGDEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
QY 241 NLIHARIEENIRAKIRRLQATEPDGCKDALQLLIEHSEMERGERLDMQALKQSTELLFG 300
Db 241 NLIHARIEQNIRAKICGLRASEAGQCKDALQLLIEHSEMERGERLDMQALKQSTELLFG 300
QY 301 GHETTASATSLITYLGLYPHVLQKVREETIKSKGLCKSNQDNKLDMEITLQLYIGCVI 360
Db 301 GHETTASATSLITYLGLYPHVLQKVREELSKGLCKSNQDNKLDMEITLQLYIGCVI 360
QY 361 KETLRNLPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADIFTNKEEFPDRFI 420
Db 361 KETLRNLPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADIFTNKEEFPDRFS 420
QY 421 VPHPEDASRFSFIPEGGLRSCVKGFEAKILLKFTVELARHCDMQLNGPPTMTSPTV 480
Db 421 APHPEDASRFSFIPEGGLRSCVKGFEAKILLKFTVELARHCDMQLNGPPTMTSPTV 480
QY 481 YPVDNLPARFTYFOGDI 497
Db 481 YPVDNLPARFTFHGEI 497
```

```
RESULT 3
US-08-882-164D-4
; Sequence 4, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
```

APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-4

Query Match 93.3%; Score 2437; DB 4; Length 497;
Best Local Similarity 93.4%; Pred. No. 6.7e-257;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

OY 1 MGLPALASALCTFVLPLLLFLAALKLMDLYCVSSRDSCALPLPPGTMGFPFEGETLQM 60
DB 1 MGLPALASALCTFVLPLLLFLAALKLMDLYCVSSRDSCALPLPPGTMGFPFEGETLQM 60
OY 61 VLQRRKFLQMKRRKYGYFYKTHLFGPRTVRVMGADNVRRIILGHRLLVSVHWPASVRTIL 120
DB 61 VLQRRKFLQMKRRKYGYFYKTHLFGPRTVRVMGADNVRRIILGDRLLVSVHWPASVRTIL 120
OY 121 GAGCLSNLHDSHKKQKRVIMQAFSREALQCYVLVIAEVSCLQWLSCGERGLVYPE 180
DB 121 GSGCLSNLHDSHKKQKRVIMRAFSEKALECYVPVITEEVGSLQWLSCGERGLVYPE 180
OY 181 VKRLMERIAMRILGCEPAGGEGDEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
DB 181 VKRLMERIAMRILGCEPQLAGDGEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
OY 241 NLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
DB 241 NLIHARIEQNIIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
OY 301 GHETTASATSLITYGLYPHYLVQKVRREIKSKGLCKSNQDNKLDMETLEQLKYICVI 360
DB 301 GHETTASATSLITYGLYPHYLVQKVRREIKSKGLCKSNQDNKLDMETLEQLKYICVI 360
OY 361 KETRLNRPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVADIFTNKEEFNPDREI 420
DB 361 KETRLNRPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNPDREI 420
OY 421 VPHPEDASRFSFIPFGGGLRSCVCKEFAKILKIFTVELARHCDWQLNGPPTMKTSPTV 480
DB 421 APHPEDASRFSFIPFGGGLRSCVCKEFAKILKIFTVELARHCDWQLNGPPTMKTSPTV 480
OY 481 YPVDNLPARFTYFOGDI 497
DB 481 YPVDNLPARFTYHGEI 497

RESULT 4
US-08-724-466B-2
; Sequence 2, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9

COUNTRY: Canada
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-2

Query Match 66.4%; Score 1734; DB 3; Length 492;
Best Local Similarity 67.7%; Pred. No. 2.9e-180;
Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

OY 1 MGLPALASALCTFVLPLLLFLAALKLMDLYCVSSRDSCALPLPPGTMGFPFEGETLQM 60
DB 1 MGLYTLMTFLCTIVLPVLLFLAALKLWEMLMIRYDPCRSPLPPGTMLPFIGETLQL 60
OY 61 VLQRRKFLQMKRRKYGYFYKTHLFGPRTVRVMGADNVRRIILGHRLLVSVHWPASVRTIL 120
DB 61 ILQRRKFLRMKROKGYCYIKTHLFGNPTVRVMGADNVRQILLGEHKLVSQWMPASVRTIL 120
OY 121 GAGCLSNLHDSHKKQKRVIMQAFSREALQCYVLVIAEVSCLQWL--SCGERGLV 177
DB 121 GSDRLSNVHGQHKKKKAIKRAFSDALEHYIPVIOQEVKSAIQEWLQKDSCLV 176
OY 178 YPEVKRLMERIAMRILGCEPAGGEGDEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGV 237
DB 177 YPEMKRLMERIAMRILGFEPEQI--KTDEQQLVEAFEEEMIKNLSLPIDVPFSGLYRGL 234
OY 238 KARNLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSWGERLDMQALKOSTEL 297
DB 235 KARNLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSWGERLDMQALKOSTEL 293
OY 298 LFGGHETTASATSLITYGLYPHYLVQKVRREIKSKGLCKSNQDNKLDMETLEQLKYIG 357
DB 294 LFGGHETTASATSLVMEFLGNTEVYQKREEVQEKVEMGYTPGKGLSMELLDQLKYTG 353
OY 358 CVIKETRLNRPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVADIFTNKEEFNPD 417
DB 354 CVIKETRLNRPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVADIFTNKEEFNPD 413
OY 418 RFIIVPHPEDASRFSFIPFGGGLRSCVCKEFAKILKIFTVELARHCDWQLNGPPTMKT 477
DB 414 RFMSKGLDEGSRFNYIPFGGGRMCVCKEFAKILKIFTVELTQHCHNWLNSNGPPTMKTG 473
OY 478 PTIVPVVDNLPARFTYF 493
DB 474 PTIYPVDNLPTKFTSY 489

RESULT 5
US-08-882-164D-2
; Sequence 2, Application US/08882164D
; Patent No. 6306624

```

;
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-882-164D-2

```

Query Match 66.4%; Score 1734; DB 4; Length 492;
Best Local Similarity 67.7%; Pred. No. 2.9e-180;
Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

```

QY 1 MGLPALASALCTFVLPVLLFLAALKLMDLYCVSSRDRSCALPLPGTMGFPPFGETLQM 60
   ||| | : ||| ||| : ||| ||| : ||| : | : ||| ||| ||| ||| :
Db 1 MGLYTLMTVLTCTIVLPVLLFLAAVKLWEMLMIRYDPCNCRSPPLPGTMGLPFIGETLQ 60
QY 61 VLQRRKFLQMKRRKYGFIYKTHLFGRPYRVMGADNVRRILLGEHRLVSVHMPASVRTIL 120
   : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| |||
Db 61 ILQRRKFLRMKQKYGCIYKTHLFGNPYRVMGADNVROILLGEHKLVSVMFASVRTIL 120
QY 121 GAGCLSNLHDSHKQRRKVIQAFSREALQCYVLVIAEVS SCLQWL---SCGERGLLV 177
   | : ||| : | : ||| ||| : ||| : | : ||| | : ||| : |||
Db 121 GSDTL SNVHG VQHKNKKAIMRAFSRDAL EHYIPV IQQEVKSAIQEWLQK DSC---VLV 176
QY 178 YPEVKRLMFRIMRILLGCEPAGGGEDEQQLVEAF EEMTRNLFSLPIDVPFSGLYRGV 237
   ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| : ||| |||
Db 177 YPEMKKLMEFRIMRILLGFEPEQI--KTDEQELVEAF EEMIKNLFSLPIDVPFSGLYRGL 234
QY 238 KARNLIHARIEENIRAKIRRLQATEPDDGCKDALQ LLIHESWGERLDMQALQKOSTEL 297
   : ||| ||| : ||| ||| : | : ||| ||| : ||| : | : ||| : |||
Db 235 RARNFIHKSIEENIRKKIQD--DDNENEQKYKDALQ LLIENSRSRSDPEPSLQAMKEATEL 293
QY 298 LFGGHETASATSLITYLGLYPHVILQKVREIKSKGL LCKSNQDNKILDMETLEQLKYIG 357
   ||| ||| ||| ||| : ||| | : ||| ||| : | : ||| ||| : ||| ||| |||
Db 294 LFGGHETASTATSLVMFLGLNTEVQKVREEVQEKV EMGMYPGKGISMELLDQLKYTG 353
QY 358 CVIKETILNPPVPGGFRVALKTFELNGYQIPKGWNV IYSICDTHDVADIFTNKEEFNPD 417
   ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : |||
Db 354 CVIKETILRINPPVPGGFRVALKTFELNGYQIPKGW NVIYSICDTHDVADVPFNKEEFQPE 413

```

```

QY 418 RFIVPHPEDASRFSFIPEGGLRSCVCGKEFAKILLKI FTVELARHCDMQLNGPPTMKT 477
   || : ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 RFMSKGLDGSFRFNYPFGGSRMCVCGKEFAKAVLLKIF LVELTQHCHWILSNGPPTMKTG 473
QY 478 PTVPVDNLPARFTYF 493
   || : ||| ||| : ||| :
Db 474 PTIYPVDNLPTKFTSY 489

```

```

RESULT 6
US-08-622-166A-2
; Sequence 2, Application US/08622166A
; Patent No. 5952545
; GENERAL INFORMATION:
; APPLICANT: KONCZ, CSABA
; APPLICANT: MATHUR, JAIDEEP
; APPLICANT: SZEKERES, MIKLOS
; APPLICANT: ALTMANN, THOMAS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
; TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESI
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,166A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION*:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 0147-0153P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-622-166A-2

```

Query Match 17.2%; Score 450; DB 2; Length 472;
Best Local Similarity 26.6%; Pred. No. 2.8e-40;
Matches 134; Conservative 90; Mismatches 217; Indels 62; Gaps 15;

```

QY 6 LLASALCTFVLPVLLFLAALKLMDLYCVSSRDRSCALPLPGTMGFPPFGETLQMV---- 61
   ||| | : ||| | : ||| ||| : ||| : | : ||| ||| ||| ||| :
Db 9 LLSIAAGFLLL-----RRTYRRMGLPPGSLGLPIGETFQ LIGAYK 52
QY 62 -LQRRKFLQMKRRKYGFIYKTHLFGRPYRVMGADNVRRILLGEHRLVSVHMPASVRTIL 120
   | : ||| : | : ||| ||| : ||| : | : ||| | : ||| : |||
Db 53 TENPEPFIDERYARYGSVFMTHLFGEPITFSADPETNR FVLQNEGKLFECSPASICNLL 112
QY 121 GAGCLSNLHDSHKQRRKVIQAFSREALQCYVLV-IAEVS SCLQWLSCGERGLLYYP 179
   | : ||| : | : ||| : | : ||| ||| : ||| : | : ||| : |||
Db 113 GKHSLLMKGSHLRMHSLLTMSFANSITIKDHLMDIDRLVRFNLDSSWSS----RVLLME 168
QY 180 EVKRLMFRIMRILLGCEPAGGGEDEQQLVEAF EEMTRNLFSLPIDVPFSGLYR-GVK 238
   | : ||| : ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : |||
Db 169 EAKKITFELTVKQLMSFDP-----GEMSESLRKEYLLVIEGFFSLPLPL-FSTTYRKA IQ 222

```


[illegible]

RESULT 7
US-08-622-166A-4
; Sequence 4, Application US/08622166A
; Patent No. 5952545

```

: GENERAL INFORMATION:
: APPLICANT: KONCZ, CSABA
: APPLICANT: MATHUR, JAIDEEP
: APPLICANT: SZEKERES, MIKLOS
: APPLICANT: ALTMANN, THOMAS
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
: TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-0747
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/622,166A
: FILING DATE: 27-MAR-1996
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: SYENSSON, LEONARD R.
: REGISTRATION NUMBER: 30,330
: REFERENCE/DOCKET NUMBER: 0147-0153P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 205-8000
: TELEFAX: (703) 205-8050
: TELEX: 248345
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 472 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-622-166A-4

```

Query Match	17.2%;	Score 450;	DB 2;	Length 472;
Best Local Similarity	26.6%;	Pred. No. 2.8e-40;		
Matches 134: Conservative	90;	Mismatches 217;	Indels 62;	Gaps 15;

```
6 ILASALCTFVLPPLLEPAALKLMDLYCVSSRRDRSCALPLPPGTMGFPEFGETLOW-----61
```

[illegible]

RESULT 8
US-08-991-677-4
: Sequence 4, Application US/08991677A

```

; Patent No. 6252135
;
; GENERAL INFORMATION:
;
; APPLICANT: Chiang, Vincent L
;
; APPLICANT: Carraway, Daniel T
;
; APPLICANT: Smeltzer, Richard H
;
; TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
;
; FILE REFERENCE: 50617
;
; CURRENT APPLICATION NUMBER: US/08/991,677A
;
; CURRENT FILING DATE: 1997-12-16
;
; EARLIER APPLICATION NUMBER: US 60/033,381
;
; EARLIER FILING DATE: 1996-12-16
;
; NUMBER OF SEQ ID NOS: 11
;
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 4
;
; LENGTH: 511
;
; TYPE: PRT
;
; ORGANISM: Liquidambar styraciflua
;
; US-08-991-677-4

```

Query Match	13.7%;	Score 358.5;	DB 4;	Length 511;
Best Local Similarity	26.7%;	Pred. No. 3e-30;		
Matches 135; Conservative	86;	Mismatches 200;	Indels 85;	Gaps 21;

QY 14 FVLPLLLFLAALKEMDLYCVSSRRDRSCALPPLPGTGMGPFEEGTILOMYLQRRKFLQMKRR 73
18 FIIPLLLGLV-----SRLRQ-RLPYPPGPKGLPVIGNMLMDQLTHRGIAKLAK 67
Db

QY 74 KYGFIYKTHLFGPTVRVMGADNVRRIILGEHRLVSVHWPAVPTLLGAGCLSNL-HDSS 132

```

Db 68 QYGGFLHLMGFLLHMVAVSTPDMARQVLQVQDNIFS-NRPATI-----AISYLYTDRA 119

```

QY 133 -----HKQRKVMQAFSRALQCY-----VVIAEVSSCLEQWLSG 171

```
Db      120 DMAFAHYGPFWRMKLCVMTLSRKRAESWESVRDEYDSAVRVASNI GST----VNIG 175
```

OY	172	EKGLLVEPVCKRLMERIAMRILLGCEPAGGEDEOQLVEAFEEMT	R--NLFSLPIDVP	229
Dd	176	E---LVFALTKNITTYRAFGTI-----SHEDQDEFVAILQEFFSOLF	GAFNIADFI	223
OY	230	F-SGLYRGV----KARNLIHARIENIRAKIRR--LOATEPDGGCKDALQ	LLEHSWER	281
Dd	224	WLKWVPOGINVRINKARGALDGFIDKIIDDHIQGSKNSEEVDTDMVDLLA	FAYGEBAKV	283
OY	282	GERLDMQ-ALK-----QSSTELLFGHETTASAATSLITYLGlyPHVLQKYREET	KSKG	334
Dd	284	SESDDLQNSIKLTKNIDIKAIMDVMEFGTETVASAIEMAMTELKMSPEDE	KKKVOE-----	338
OY	335	LLCKSNQDNKLDMETLEQKVICVAKETLURLNPYPVPGFRVALKTFFELNGYOI	PKGMNV	394
Dd	339	LAVVGGLDRVREEKDFEKLYLKCVLKEVLRHLHPPIPLLHETAEDA	EYGVGIIPAKSRV	398
OY	395	IYSICDTHDVADIFTNKKEEFNPDRTIVPHPED--ASRFSFIPFGGLRSVCYKEFAKIL	L	452
Dd	399	MINACAIGRDKNSWADPDTRFPRFKLDGVDPDKGNNEFIFPGSGRRSCPQMOLGYAL		458
OY	453	KIFTVELARHC-DWOLLNGPPIMKTS	477	
Dd	459	ET-TVAHLHCFTWEL--PDGMKPS	480	

RESULT 9
 US-09-091-432-2
 Sequence 2, Application US/09091432
 Patent No. 5981837
 GENERAL INFORMATION:
 APPLICANT: Chapple, Clint
 TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition
 FILE REFERENCE: 7024-325
 CURRENT APPLICATION NUMBER: US/09/091,432
 CURRENT FILING DATE: 1998-06-18
 EARLIER APPLICATION NUMBER: PCT/US96/20094
 EARLIER FILING DATE: 1996-12-19
 EARLIER APPLICATION NUMBER: US 60/009,119
 EARLIER FILING DATE: 1995-12-22
 EARLIER APPLICATION NUMBER: US 60/013,388
 EARLIER FILING DATE: 1996-03-14
 NUMBER OF SEQ ID NOS: 3
 SOFTWARE: Microsoft Word 2.0C
 SEQ ID NO 2
 LENGTH: 520
 TYPE: PRT
 ORGANISM: Artificial
 FEATURE:
 OTHER INFORMATION: Sequence is deduced from DNA sequence of SEQ ID NO:1
 Patent No. 5981837
 US-09-091-432-2

[illegible][illegible]

```

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
104
```

```

Query Match          11.5%; Score 300; DB 3; Length 576;
Best Local Similarity 23.7%; Pred. No. 8.7e-24;
Matches 128; Conservative 86; Mismatches 187; Indels 138; Gaps 21;

QY 41 ALPLPGT----MGFPFGETLQWVLQRRKFLQMKRRKRYGFYKTHLFGKPTVRVMGADN 96
   ::::: | ::::: | ::::: | ::::: | ::::: | ::::: | ::::: |
Db 78 SMPiEAGAVSDLLGRPLDFSLYDWFLE-----HGAVYKL-AFGPKAFVWVSDPI 125

```

[illegible]

```

RESULT 11
US-08-948-564-4
; Sequence 4, Application US/08948564
; Patent No. 6121512
;
; GENERAL INFORMATION:
; APPLICANT: Siminszky, Balazs
; APPLICANT: Dewey, Ralph E.
; APPLICANT: Corbin, Frederick T.
; TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
; NUMBER OF SEQUENCES: 23
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 6121512th Carolina
; COUNTRY: USA
; ZIP: 27627
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,564
; FILING DATE:
;
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 5051-409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 510 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein

```

US-08-948-564-4

Query Match	10.9%;	Score 284.5;	DB 3;	Length 510;
Best Local Similarity	24.4%;	Pred. No. 3.5e-22;		
Matches 117; Conservative	93;	Mismatches 222;	Indels 47;	Gaps 16;

QY	18	L L L F L A L K L M D L Y C V S S R D R S C A L B L P R G T M G F P F E G L I O W -	-L Ö R K K F L O M K R R K Y	75
D b	20	L F I F F V F E F K L Y Ö R - -	S D S K T S S T C K L P P G P R T L P L I G N I H Q I V G S L P V H Y L K N L A D K Y	76
QY	76	G F I Y K T H L - F G R P T V R V M G A D N V R R I L L G E H R L V S V H M P A S V R T I L - - -	G A G C I S N L H D	130
D b	77	G P L - - M H L K L E G V S N I I V T S P E M A Q E I M K T H D L N F S D R P D F V L S R I V S Y N G S G I V F S Q H G		134
QY	131	S S H K Ö R K K V - I N O A F S R E A L Ö C Y V L Y I A E E V S S C L E Ö - W L S C G E R G L L V - - -	P E Y K R I M	185
D b	135	D Y W R Ö L R K I C T V E L L T A K R V Ö S F R S I R E E E V A E L V K K I A A T A S E E G G S I F N L T Ö S I Y S M T		194
QY	186	F R I A M R I L L G C E P G P A G G E D E Q Ö L V E A F E E M T R N L - F S L P I D V P F S G L Y R G V K A R N L I		243
D b	195	F G I A A R A F G K K - - - - - S R Y Ö Q V F I S M H K Ö L M L L G F S V A D L Y P S S R V F Ö M M G A T G K L		248
QY	244	H A - - R I E E N I R A K I - - - - - R L Ö A T E P R D G C K D A L Ö L L I E H S W E R G E R L D M A L K Ö S T E		296
D b	249	E K V H R V T D R V L Ö D I I D E H K N R N R S S E E R A V E D L V D L L K F Ö K S E F R L T D N I K A V I Ö D		308
QY	297	L L F G H E T T A S A T S L I T Y L G L Y P H V L Ö K Y R E I K - - - S K U L L C K S N Ö D N K I D M E T L E Q		352
D b	309	I F I G G E T S S S V E M G M S E L I R N P R V M E E A Ö A E V R R V Y D S K G Y - - - - - V D E T E L H Q		359
QY	353	L K Y I G C V I K E T L R L N P P V P G G F - R V A L K T F E L N G Y Ö I P K G W N V I Y S I C D T H V A D I F T N K		411
D b	360	L Y L K S I I K E T M R L H P P V L L V P R V S R E R Q I N G Y E I P S K T R I I N A M A I G R N P K Y W G E T		419
QY	412	E E F N P D R F I V P H P E - D A S R F S E I P F G G L R S C V G K E F A K I L K I T Y E L A R H C D W Ö L N		469
D b	420	E S K R P E R L N S S I D E R G T D E F I P E G A G R I C I T P I T A I P N I E L P L A Ö L L Y H F D W K L P N		478

RESULT 12
US-08-313-075A-50
Sequence 50, Application US/08313075A
Patent No. 5639870
GENERAL INFORMATION:
APPLICANT: Holton, Timothy A.
APPLICANT: Cornish, Edwina C.
APPLICANT: Tanaka, Yoshikazu
TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID
TITLE OF INVENTION: PATHWAY ENZYMES AND USES THEREFOR
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,075A
FILING DATE: 30-NOV-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL 1538/92
FILING DATE: 27-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL 6698/93
FILING DATE: 07-JAN-1993

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PCT/AU93/00127
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digilio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 9433
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-313-075A-50

```

Query Match 10.4%; Score 270.5; DB 1; Length 496;
Best Local Similarity 23.2%; Pred. No. 1.1e-20;
Matches 117; Conservative 66; Mismatches 197; Indels 125; Gaps 16;

```

QY 33 VSSRDSCALPLPPTMGFPFEGTLQWLVQRRKFLQMKRRKYGYTKTHLFGPTVRVM 92
   :| | | : | | | | : | | : | | : | | :
Db 21 MSLRKRKISKLPKPPPLPIGHLGNHPKSLAQAKIHGPIMLNLKLGQINTVIS 80

QY 93 GADNVRRL-----LGEHRLVSVHM-PA-----SVRTILGAGCLS--N 127
   : | : | : | : | : | | | : | | : | | :
Db 81 SSVVAREVLQKODLTFSNRFVPDVVHVNHSDFSVVWLPVNSRWKTLRKIMNSSIFSGNK 140

QY 128 LHDSSHKQKKVIMQAFSREALQCYLVIAEEVSSCLEQWLSGGERGLLVYPEVKRLMR 187
   | : | : | | | : | : | : | : | : | : |
Db 141 LDGNQHLRSKVV-----QELIDYCK---CAKNGEAV--DIGRATEG 177

QY 188 IAMRLLGCEPGPAGGDEQQLVEAFEEMTRNLFSLPIDVPFSGLYR----- 235
   : : | : | : | : | : | : | : | : | :
Db 178 TTLNLL-----SNTIFSKDLTNPFSDSAKEFEKELVWNIMVE 213

QY 236 -----GVKAR-----NLIHARIEENIRAKIRRLQATEPDGCG 267
   : | : | : | : | : | : | : | : | : |
Db 214 AGKPNLVDPFPLEKIDPGIKRRMTNNFTKFLGLISGLIDD--RLKERNLR---DNAN 267

QY 268 KDALQLLIEHSWGERLDMQALQKOSTELLFGHETTSAAATSLITYLGLYPHYLQYR 327
   | : | : | | | : | : | : | : | : | : |
Db 268 IDVLDAALLNISQENPEEIDRNOIEQLCLDLFAAGTDTSTLEWAMAEILLQNPHTLQRAQ 327

QY 328 EEIKSKGLCKSNQDNKLDMETLEQLKYIGCVIKETLRPNPVPVGGF--RVALKTFEELNGY 386
   || : : | : | : | : | : | : | : | : |
Db 328 EELAQ--VIGKGQVEAD---VGRPLRLCIYKETLRIRHPAAPLLIPRKEVEDVELSTY 382

QY 387 QIPKGMNVIYSICDTHVDADIFTNKEEFPNDRFIVHPEDASR--FSFIPIFGGLRSCVGR 445
   || | : : : | : | : | : | : | : | : |
Db 383 IIPKDSQVLVNWVAIGRNSDLMENPLVEKPERFEWESEIDIRGRDFELIPFAGRRICPGL 442

QY 446 EFAKILKIFTVELARHCDWQLNG 470
   | : : : | : | : | : | : | : | : |
Db 443 PLAMRMIPVALGSLNSFNWKLKYG 467

```

```

RESULT 13
US-09-292-768-68
; Sequence 68, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; TITLE OF INVENTION: LIMONENE HYDROXYLASES
; FILE REFERENCE: wsur13463
; CURRENT APPLICATION NUMBER: US/09/292,768

```

```

; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 68
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Artificial Sequence
; US-09-292-768-68

```

Query Match 10.0%; Score 262.5; DB 4; Length 500;
Best Local Similarity 21.9%; Pred. No. 8.4e-20;
Matches 115; Conservative 89; Mismatches 215; Indels 105; Gaps 19;

```

QY 7 LASALCTFVLPILLFLAALKLMDLYCVSSRDSCALPLPPTMGFPFEGTLQWLVQ-- 64
   :| | : | : | : | : | : | : | : | : |
Db 5 ISSAIIILVYTYTISLLIIKQW-----RKPQENLPKPPKPLIGH-LHLWGKLP 56

QY 65 RKFLQMKRRKYGYTKTHL-----FGRPTVRVMGADN 96
   : | : | : | : | : | : | : | : | : |
Db 57 QHALASVAKQYPAHVQOLGEVFSVLSREATKAMKLVDPACADRFESIGTKIMWYDN 116

QY 97 VRRILGHEHRLVSVHWPASVRTILGAGCLSNLHDSHKQKKVIMQAFSREALQCYLVY 156
   | : | : | | | : | : | : | : | : | : |
Db 117 -DIIIFSPY---SVHW-----RQMRKICVSELLSARNVRSFGFIR 152

QY 157 AEEVSSCLEQWLSGGERGLLVYPEVKRIMERIA-MRILLGCEPGAGGDEQQLVEAFE 215
   :| | | : | | : | : | : | : | : | : |
Db 153 QDEVSRLLGHLRSSAAG-----EAVDLTERIATLTCSIIICRAAFGSVIRDHELV---- 203

QY 216 EMTRNLFSLPIDVPFSGLYRGVKARNLI-----HARIEENIRAKI--RRLQATEP 263
   | : : | : | : | : | : | : | : | : |
Db 204 ELVKDALSMASGEFELADMFPSSKLNLMLCWNSKLMWRMRRRVDAIIEAIVEEHKIKSGE 263

QY 264 DGGCDALQLLIEHSWGERLDM--QALKOSTELLFGHETTSAAATSLITYLGLYPH 321
   || : : | : | : | : | : | : | : | : |
Db 264 FGG-EDIIIDVLFMRQKDSQIKVPTTKAIFKAFIDTFSAGETFSSTTLWVMAELMRNPE 322

QY 322 VLQKVBREIKSKGLCKSNQDNKLDMETLEQLKYIGCVIKETLRPNPVPVGGFVALKTF 381
   | : | : | : | : | : | : | : | : | : |
Db 323 VMAKQAQEVRA--ALKGKTD--WDVDDVQELKYMKSVMKETMRMHPPIPLIPRSCREEC 377

QY 382 ELNGYQIPKGMNVIYSICDTHVDADIFTNKEEFPNDRFIVHPEDASR-----FSFIPIF 435
   | : | | | | : : : | : | : | : | : |
Db 378 EVNGYTIIPNKARIMINWMSGRNPLWYKEPETFWPERF-----DQVSRDFMGNDFFETP 432

QY 436 GGGLRSCVGEFAKILKIFTVELARHCDWQLNG--PPTMKTS 477
   | | | | | | : : : | : | : | : | : |
Db 433 GAGRRICPGLNFGLANVEVPLAQLLYHFDWKLAEGMNPSPMDMS 476

```

```

RESULT 14
US-09-292-768-4
; Sequence 4, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; TITLE OF INVENTION: LIMONENE HYDROXYLASES
; FILE REFERENCE: wsur13463
; CURRENT APPLICATION NUMBER: US/09/292,768
; EARLIER FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 500
; TYPE: PRT

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:35:30 ; Search time 71.14 Seconds
(without alignments)
532.173 Million cell updates/sec

Title: US-09-668-482-32
Perfect score: 2612
Sequence: 1 MGLPALLASALCTFVLPILL.....PTVYPVDNLPARTYFQGDI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	668.5	25.6	444	1	S75761	cytochrome P450 -
2	511.5	19.6	485	2	A84859	probable cytochrom
3	501.5	19.2	482	2	T02739	probable cytochrom
4	494.5	18.9	457	2	T04444	cytochrome P450 -
5	461	17.6	513	2	T46143	steroid 22-alpha-h
6	450	17.2	472	1	S55379	cytochrome P450 cy
7	442	16.9	464	2	T07859	cytochrome P450 ho
8	442	16.9	490	2	H86185	hypothetical prote
9	418	16.0	452	2	D85429	cytochrome P450 11
10	410	15.7	492	1	D70649	cytochrome P450 Ry
11	406	15.5	489	2	B84733	probable cytochrom
12	359	13.7	519	1	T02263	cytochrome P450 DW
13	339	13.0	486	2	JC4240	lanosterol 14alpha
14	339	13.0	503	2	JC4758	lanosterol 14alpha
15	337	12.9	520	2	T04591	ferulate-5-hydroxy
16	332	12.7	503	2	JC7243	sterol 14alpha-dem
17	329.5	12.6	512	2	H96759	probable steroid 2
18	326	12.5	503	2	JC4759	sterol 14-demethyl
19	325	12.4	511	1	A39381	cytochrome P450 4
20	323.5	12.4	496	1	S41192	cytochrome P450 4D
21	323.5	12.4	511	1	A34260	laurate omega-hydr
22	323	12.4	472	1	G70932	probable monoxigen
23	323	12.4	509	1	S68855	lanosterol 14alpha
24	322.5	12.3	476	1	F70791	cytochrome P450 Ry
25	317	12.1	441	1	C70616	cytochrome P450 Ry
26	315.5	12.1	495	2	T06288	probable cytochrom
27	314.5	12.0	512	2	T13611	hypothetical prote
28	312.5	12.0	520	2	H84663	probable cytochrom
29	310	11.9	509	2	S47553	cytochrome P450 Cy

30	308.5	11.8	519	2	I53015	fatly acid omega-h
31	308.5	11.8	591	2	I65981	fatly acid omega-h
32	307	11.8	511	1	S41598	cytochrome P450 77
33	305.5	11.7	495	2	T52515	related to n-alkan
34	305.5	11.7	519	2	JX0331	laurate omega-hydr
35	305	11.7	503	2	T24985	hypothetical prote
36	302	11.6	511	1	S25707	cytochrome P450 en
37	302	11.6	513	2	JC7120	cytochrome P450 4A
38	301.5	11.5	511	1	B34160	cytochrome P450 4A
39	301	11.5	520	1	A46661	leukotriene B4 ome
40	300.5	11.5	537	2	JC4534	cytochrome P450 4F
41	300	11.5	576	2	T05904	cytochrome P450 97
42	299.5	11.5	509	1	O4RTLO	laurate omega-hydr
43	296.5	11.4	382	2	T48613	hypothetical prote
44	296.5	11.4	510	1	A34160	laurate omega-hydr
45	292	11.2	510	1	A29368	prostaglandin ome

ALIGNMENTS

RESULT 1	
S75761	cytochrome P450 - Synechocystis sp. (strain PCC 6803)
N;Alternate names: protein slr0574	
N;Contains: oxidoreductase (EC 1.---.---)	
C;Species: Synechocystis sp.	
A;Variety: PCC 6803	
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000	
C;Accession: S75761	
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O., K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas	
DNA Res. 3, 109-136, 1996	
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys	
S.	
A;Reference number: S74322; MUID:97061201	
A;Accession: S75761	
A;Status: nucleic acid sequence not shown; translation not shown	
A;Molecule type: DNA	
A;Residues: 1-444 <KAN>	
A;Cross-references: EMBL:D64003; GB:AB001339; NID:g1001200; PIDN:BA10496.1; PID:g100	
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996	
C;Genetics:	
A;Gene: cyp	
C;Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology	
C;Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase	
F;251-413/Domain: cytochrome P450 homology <P45>	
F;391/Binding site: heme iron (Cys) (axial ligand) #status predicted	
Query Match	25.6%; Score 668.5; DB 1; Length 444;
Best local Similarity	34.1%; Pred. No. 1.7e-44;
Matches 154; Conservative 87; Mismatches 181; Indels 29; Gaps 8;	
OY 41 ALPLPGTMGPFPEFGETLQMVLRKRLQMKRRKYGFIYKTHLFGPTVRVMGADNVRRI 100	
Db 10 SLPIPPGDFGLPWLGETLNF-LNDGDFGKKRQOQFGPIFKTRLFGKNVIFISGALANRFL 68	
OY 101 LIGEHRLVSWHPASVRTILGAGCLSNLHDSHKORKKVTIMQAFSREALQCYVLVIAEEV 160	
Db 69 FTKEQETFGATWPLSTRILLGNALATQMGELIHSRRKILYQAFLLPRTLDSTYLPKMGITV 128	
OY 161 SSCLQWLSCGREGLLVPEVKRIMRIAMRILLGCEPAGGGEDEQOLVEAFEEWTRN 220	
Db 129 QGYLEQWGRKANE--VIWYPQLRMTFDVAATLFMGEKV-----SQNPQLFPWFETVIOG 180	
OY 221 LFSLPIDVPSGLYRGVKARNLIHARIEENIRAKIRRLQATEPDDGCKDALQLIHSWE 280	
Db 181 LFSLPIDVPSGLYRGVKARNLIHARIEENIRAKIRRLQATEPDDGCKDALQLIHSWE 234	
OY 281 RGERLDMQALKQSSTELLFGHETTASATSLITYLGLYPHVLOKVREIKSGILCKSN 340	
Db 235 NNQPLSLPELKDQILLLLFAGHETTLTALSFFCLLLGQHSIDIRERVQEQN-----KLQ 288	

```

OY      341  ODNKIDMETLEQLKYIGCVIKETIRLNPPVPGGFVALKTFELNGYQIPKGMNVIYSICD  400
        :| |||::: | :|: || || ||| ||| : : | : ||| | | |
Db      289  LSGELTAETLKMPYLDQVLQEVRLRIPVGGGFRELQDCQGFHFHPKGLVSYQISQ  348
OY      401  THDVADIFTNKEEFNPDRFIVHPEDASR---FSFIPEGGLRSCVYGKEFAKILKIIFT  456
        || : : : : | : | : | | : : : : | : ||||| | : ||| : : | : |
Db      349  THADPDLYPDPEKFDPERFT---PDGSATHNPPFAHVHPFGGLRECLGKEFARLEMKLFA  405
OY      457  VELARHCDWQLNGPP-TMKTSPTYVYPVDNL  486
        | : |||| | : : : | |||
Db      406  TRLIQGFQDWTLIPGQNLVELVTPSPRPKDNL  436

```

RESULT 2
 A:Accession: A84859
 Probable cytochrome P450 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 23-Mar-2001
 C:Accession: A84859
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: A84859
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-485 <STO>
 A:Cross-references: GB:AE002093; NID:g4512670; PIDN:AAD21724.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g42850
 A:Map position: 2
 C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: heme; iron; metalloprotein
 F:432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	19.6%;	Score 511.5;	DB 2;	Length 485;
Best Local Similarity	28.3%;	Pred. No. 3.6e-32;		
Matches 132;	Conservative 86;	Mismatches 201;	Indels 47;	Gaps 10;

```

QY 44 LPPGTMGEFFEGETLQWVLQRRK-----FLQMKRRKYGEIYKTHLFGRPVRYMGADNV 97
      | | | | | : | | : : : : : : : | : | : | : | : | : | : | :
Db 45 LLPGEMGLPWIGETMDYKAQKSNRVEDEFPNPRIKHGNIEKTRIMGSPFIYVNGAEAN 104
      | | | | | : | | : : : : : : : | : | : | : | : | : | : | :
QY 98 RRLIGEHLVSVHWPASVFTILGAGCLSNLHDSHQRKKVIMQAFSREALQCVLVIA 157
      | | | | | : | | : : | : : | : : : : : : : | : | : | : | :
Db 105 RLILSNEFSLVSSWSSSSVQLMGMCIMAKQGEKHRVLRGIVANSLSYIGLESLLPKLC 164
      | | | | | : | | : : | : : | : : : : : : : | : | : | : | :
QY 158 EEVSSCLE-QWLSGGERGLLVPEVKRLMFRIAMRILGCEPAGGEGEEOOLVEAFEE 216
      | : | : | : : : : : | : | : | : | : | : | : | :
Db 165 DTVKFHHEETWR--GKEEISLYRSAKAVLTFYVFECLYGIVK-----EIGMLEVFER 214
      | : | : | : : : : : | : | : | : | : | : | : | :
QY 217 MTRNLFSLPIDVPFSGLYRGVKARNLIHARIENIRAKIRRLQATEPDGCKDALQLLIE 276
      : : | : | : | : | : | : | : | : | : | : | : | : | :
Db 215 VLEGVFALPVEFPCKSFARAKKARLEIETFLVGKVRERKREM--EKEGAEKPNNTLLF-- 269
      | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 277 HSWERGERLDMQALKOSTE-----LLFGHETTASATSLITYLGLYPHVLQKVR 327
      | | : : : : | : | : | : | : | : | : | : | : | : | :
Db 270 -----SRLVEELIKGVITEEEVVDNMVLLVFAAHDTSYAMSMFVKMLAQHP----TCR 319
      | | : : : : | : | : | : | : | : | : | : | : | : | :
QY 328 EEIKSKGLICKSN--QDNKLDMETLEQLKYGIVIKETLRLNPPVPGGFRVALKTFELNG 385
      : : : : | : | : | : : : | : : | : | : | : | : | : | :
Db 320 DTLLOEHAQIKANKGEGEYLTVEDVKMKMYSQVAVRETMRLSPIFGSFRKAVADIDYGG 379
      | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 386 YQIPKGMWNVYSICDTHDVADIFTNKEFPNDRFIVPHPEDASRESFIPFGGLRSCVVGK 445
      | | | | | : : : | : : : | : | : | : : : | : | : | : | :
Db 380 YTI PKGWKITWTTGYGTHYNPEIFQDPMSFDPTRFDKP----IQAVTYLLPFGGGPRLCAGH 435
      | | | | | : : : | : : : | : | : | : : : | : | : | : | :
QY 446 EFAKILLKIFTELARHCDWQLLNGPPTMKTSPTVYPVDNLPAFFT 491
      : | | | : : | : | : | : | : | : | : | : | : | : | :

```

Db 436 Q LAKISILVEMHFVVTGFDWSLVYPDETISMDPLPPPSLGMPIKIS 481

RESULT 3

probable cytochrome P450 At2g29090 [imported] - Arabidopsis thaliana
N/Alternate names: cytochrome P450 homolog T9I4.17
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C/Accession: T02739; D84692
R/Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, August 1998
A/Description: Arabidopsis thaliana chromosome II BAC T9I4 genomic sequence.
A/Reference number: Z14710
A/Accession: T02739
A/Status: translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-482 <ROU>
A/Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461849
A/Experimental source: cultivar Columbia
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
Muss, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487
A/Accession: D84692
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-482 <STO>
A/Cross-references: GB:AE002093; NID:g3461849; PIDN:AAC33235.1; GSPDB:GN00139
C/Genetics:
A/Gene: At2g29090; T9I4.17
A/Map position: 2
A/Introns: 80/2; 187/3; 238/3; 321/3; 351/3; 413/3; 453/2
C/Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
F:291-453/Domain: cytochrome P450 homology <P45>

Query Match	19.2%;	Score 501.5;	DB 2;	Length 482;
Best Local Similarity	28.2%;	Pred. No. 2.1e-31;		
Matches 140;	Conservative 94;	Mismatches 219;	Indels 43;	Gaps 13;

```
QY      4 PALLASALCTFEVLPLLLFLAALKLMDLYCVSSRDRSCALPLPBGTMGEFFEGETLQMVLQ 63
        ||| : | : | | | : | : | | | | | | | | | | : :
Db     20 PALITLTIVVVVLLF-----KMWLHWKEQR-----LRLPPSGMGLPYIGETLRLYTE 68
        ||| : | : | | | : | : | | | | | | | | | | : :
QY      64 R-RKFLQMKRRKYGFYKTKHLFGRPYVRVMGADNVRRIILGHEHRLVSVHWPASVRTILGA 122
        | : : | | | : | | | : | : | : | : | : | : | :
Db     69 NPNSFEATRQNKYGDIKFTHILGCFCVMISSPEARMVLVSKAHLFKPTYPPSKERMICP 128
        | : : | | | : | | | : | : | : | : | : | : | :
QY     123 GCLSNLDSSHQKKVKVIMQAFSEALQCYLVIAEEVSSCLEOMWLSGGERGLVYPEVK 182
        | | : : | : : | : | : | : | : | : | : | : | :
Db     129 EALFPHQGPYHSTLKRLVOSSMPBSALRPVTSHIELVLQTLSSWTSQSINTLEY - MK 186
        | | : : | : : | : | : | : | : | : | : | : | :
QY     183 RLMEFIAMRILLGCEPGPAGGEDQQLVEA---FEEMTRNLFSLPIDVPFSGLYRGVK 238
        | | : : | : : | : | : : | : : | : | : | : | :
Db     187 RYAFDVAIM-----SAFGDKPEPTTIDVIKLLYQRLERYNSMPLDPGLTFHKSMK 238
        | | : : | : : | : | : : | : : | : | : | : | :
QY     239 ARNLHARIEENIRAKI-RLQATEPDGGCKDALQLLIEHSMERGERLDMQALKOSSTEL 297
        || | : | : | : | : | : | : | : | : | : | :
Db     239 AR-----IELSEELRKVIEKRRENGREEG--LLGVLLGAKDQKRNGLSDSQIADNIIGV 291
        || | : | : | : | : | : | : | : | : | : | :
QY     298 LEFGHETTASATSLITYLGYPHYLOKV-REEIKSKEGLICKSNODNKLDMETLEOLKIYI 356
        : | : | | | | : | : : : | : | | : : | : | :
Db     292 IFAATDTASVLTWLTKYLHDHPNLLOEVSREQFSIROKIKEKNR--RISWEDTRRMPLT 349
        : | : | | | | | : | : : | : | | : : | : | :
QY     357 GCVIKETLRLNPVPGCFRYVALKTFFELNGYOIPKGWNVIYSICDTHVDADIFTNKEEFNP 416
        ||| : | : | : | : | : | : | : | : | : | :
Db     350 TRVIQETLRAASVLSFTFEREAQDVEYDGyllPKGMXYLPLFRRIHSSSEFFPDEKEDP 409
        ||| : | : | : | : | : | : | : | : | : | :
QY     417 DRE-IVPHPEDASRFSLPFGGGLRSCVCGKEFAKILKIETFVELARHCWMQNLNGPPTMK 475
        || : | | : : : | | | | : | | | : | | | : : :
```


Db 410 SRFEVAPKP-----YTYMPFGNGVHSCPGSELAKLEMLILHLHLLTSFRWEVIGDEEGIQ 464

OY 476 TSPTVYPVDNLPARFT 491

Db 465 YGPFVPYKKGLPIRVT 480

RESULT 4

T04444

cytochrome P450 - Arabidopsis thaliana

N:Alternate names: protein T18B16.200; protein T5K18.10

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2000

C:Accession: T04444; T05806

R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansorge, W.; Bancroft, I.; Mewes, H.W. submitted to the Protein Sequence Database, April 1998

A:Reference number: Z15359

A:Accession: T04444

A:Molecule type: DNA

A:Residues: 1-457 <BEV>

A:Cross-references: EMBL:AL021687

A:Experimental source: cultivar Columbia; BAC clone T18B16

R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voelt, M.; Robben, J.; Volckaert, G.; Ba submitted to the Protein Sequence Database, April 1998

A:Reference number: Z15453

A:Accession: T05806

A:Molecule type: DNA

A:Residues: 131-457 <BEW>

A:Cross-references: EMBL:AL022580

A:Experimental source: cultivar Columbia; BAC clone T5K18

C:Genetics:

A:Map position: 4

A:Introns: 67/2; 173/3; 302/3; 358/1; 393/3

A>Note: T18B16.200; T5K18.10

C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase

F:272-433/Domain: cytochrome P450 homology <P45>

Query Match 18.9%; Score 494.5; DB 2; Length 457;

Best Local Similarity 29.5%; Pred. No. 7e-31;

Matches 137; Conservative 92; Mismatches 191; Indels 45; Gaps 13;

OY 18 LLLFLAALKLMDLYCVSSRDR--SCALPLPGTMGFPFFGETLQMVLO-RKKFLQMKRRK 74

Db 8 LTLFAGSLFLYFLRLISQRRFGSSKPLPLPGTMGMPYVGETFOLYSQDPNVFFQSKOKR 67

OY 75 YGFIYKTHLFGPTRYVMGADNVRRIILGHEHRLVSVHWPASVRTLIGAGCLSNLHDSHK 134

Db 68 YGSVEKTHVLGCPVMISSPEAKFVLVTKSHLFKPTFPASKERMLGKQATFPHQGDYHA 127

OY 135 QRRKVIQAFSREALQCYLVIAEEVSSCLEQWLSCGERGLLV--YPEVKRLMERIAMRI 192

Db 128 KLRKLVLRAPMPEISIRNVPDIESIAQDSLRSW-----EGTMINTYQEMKTYTFNVALLS 182

OY 193 LLGCEPFPAGGEDE---QOLVEAFEEEMTRNLFSLPIDVPFSGLYRGVKARNLIHARIE 248

Db 183 IF-----GKDEVLYREDLKRCCYILEKGYNSMPVNLPGTLFHKSMKARKEL-SQIL 232

OY 249 ENIRAKIRRLQATEPDDGCKDALQLLIEHSWERGERLDMQALKOSTELLFGGHETTASA 308

Db 233 ARILSERRQNGSSSHND-----LLGSFMGDKKEELTDEO-IADNIIGVIFAARDTASV 283

OY 309 ATSLITYLGLYPHYLQKVREETKSKGLLCKSNQDNKLDMETLEQLKYIGCVIKETLRNP 368

Db 284 MSWILKYLAEHPNVLEAVTEE--QMAIRKDEEGESLTWGDTRKMPPLTSRVIQETLRVAS 341

OY 369 PVPGGFRVALKTFELNGYQIPKGNVVIYSICDTHDVADIFTNKEEFNPDRE-IVPHPEDA 427

Db 342 ILSFTFREAVEDEVYEGYLIPKGVKYLPLFRNIHSHADIFSNDGKFPDPSRFEVAPKPN-- 399

OY 428 SRFSFIPEGGLRSCVGKEFAKILIKI---FTVELARHCDWQLL 468

Db 400 ---TFMPFGNGHSCPGNELAKLEMSIMTHHLTTKYRCVCVYYLL 441

RESULT 5

T46143

steroid 22-alpha-hydroxylase (DMF4) - Arabidopsis thaliana

N:Alternate names: protein T3A5.40

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 16-Feb-2001

C:Accession: T46143

R:Blöcker, H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.; Quetier, F.; S submitted to the Protein Sequence Database, December 1999

A:Reference number: Z23024

A:Accession: T46143

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-513 <BLO>

A:Cross-references: EMBL:AL132979

A:Experimental source: cultivar Columbia; BAC clone T3A5

C:Genetics:

A:Map position: 3

A:Introns: 74/2; 182/3; 233/3; 338/3; 369/3; 396/1; 432/3

A>Note: T3A5.40

C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: heme; iron; metalloprotein

F:308-484/Domain: cytochrome P450 homology <P45>

F:462/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 17.6%; Score 461; DB 2; Length 513;

Best Local Similarity 28.4%; Pred. No. 3.4e-28;

Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

OY 3 LPLALASALCTFVLPPLLLFLAALKLMDLYCVSSRDRSCALPLPGTMGFPFFGETL----- 58

Db 10 LPLLLPSL----LSLLFLILLK-----RRNRKTRFNLPPGKSGWPFLETIGYLK 57

OY 59 -QMVLORRKFLQMKRRKYGFITYKTHLFGPTRYVMGADNVRRIILGHEHRLVSVHWPASVR 117

Db 58 PYTATTLGDFMQQHVSKYGYKIRSNLFGEPITVSADAGLNFILQNEGRLLFECSPRSIG 117

OY 118 TLIGAGCLSNLHDSHKQRRKVIQAFSREALQCYLV-IAEEVSSCLEQWLSCGERGLL 176

Db 118 GILGKWSMLVYVGDMDRMSISLNFSLHARLRTLLKDVERRHTLFLVDSWQO--NSIFS 175

OY 177 YPEVKRLMERIAMRILLGCEPFPAGGEDEQOLVEAFEEEMTRNLFSLPIDVPFSGLYRG 236

Db 176 AQDEAKKTFENLMAKHIMSDPGE---EETEQLKKEYVTFMKGVSAPLNLPGTAYHRA 231

OY 237 VKARNLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSWE-----RGERLDMQ- 288

Db 232 LQSRATILKFIERKME--RKLDIKEED--QEEEVKTEDEAEEMSKSDHVRKQRTDDL 286

OY 289 ---ALKQ--STE-----LLEFGHETTASAATSLITYLGLYPHYLQKVREIYSKGL 335

Db 287 LGWVLKHSNLSTEQILLDLILSLFAGHETSSVAIALAIFFLQACPKRAVEELREHLEIAR 346

OY 336 LCKSNQDNKLDMETLEQLKYIGCVIKETLRNPVPGGFRVALKTFELNGYQIPKGNVI 395

Db 347 AKKELGESELNWDYKKMDFTQCYNETLRIGNVYRFLHRKALKDVARYYKGYDIPSGWKL 406

OY 396 YSICDTHDVADIFTNKEEFNPDREIVPHPEASRFS-----FIPFGGLRSCVGKE 446

Db 407 PVISAVHLDNSRYDQPNLFPWRWQOQNGASSSGSFSFTWGNNNYMPFGGPRLCAGSE 466

OY 447 FAKILKIFTVELARHCDWQLLNGPPTMKTSPTVYPVDNLP 487

Db 467 LAKLEMAVFIHHLVLKRWELAE-----DDKPAFPFVDFP 502

RESULT 6

S55379

Cytochrome P450 CYP90 - Arabidopsis thaliana

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C/Accession: S55379

R:Sequences, M.; Nemeth, K.; Koncz, Z.; Nagy, F.; Koncz, C.

Submitted to the EMBL Data Library, May 1995

A:Reference number: S55379

A/Accession: S55379

A:Molecule type: mRNA

A:Residues: 1-472 <SZE>

A:Cross-references: EMBL:X87367; NID:g853718; PIDN:CAA60793.1; PID:g853719

C:Genetics:

A:Gene: CYP90

C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;

F;275-440/Domain: cytochrome P450 homology <P45>

F;418/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 17.2%; Score 450; DB 1; Length 472;

Best Local Similarity 26.6%; Pred. No. 2.2e-27;

Matches 134; Conservative 90; Mismatches 217; Indels 62; Gaps 15;

QY 6 LLASALCTFVPLLLFLALKLMDLYCVSSDRSCALPLPPGTMGFFFGETLQMV--- 61

DB 9 LLSIAAGFLLL-----RRTRYRMGLPPGSLGLPLIGETFOIGAYK 52

QY 62 -LQRRKFLQMKRRKYGFYKTHLFGRPYRVWGANVRRIILGHRVSVHWPASVRTIL 120

DB 53 TENPEPFDERVARVGVFMTHLFGEPFITSADPETNRFVLQNEKLFECSPASICNL 112

QY 121 GAGCLSLNHDSSHQRRKQVIMQAFSREALQCVLV-IAEVSSCLEQWLSCGERGLLYP 179

DB 113 GKHSLLMKGSLKRMHSLTMSFANSSIIKDHMLDIDLVRNLDSSWSS---RVLLME 168

QY 180 EVKRLMFRIAMRILGCEPAGGEGEQLVEAFEEEMTRNLFSLPIDVPFSGLYR-GVK 238

DB 169 EAKKITFELTVKQLMSPD-----GEWSESLRKEYLLVIEGFFSLPL-LFSTTYRKAIO 222

QY 239 ARNLHARIEENIRAKIRRLQATEPDGG--CKDALQLLI--EHSWGERGLDMQALKOSS 294

DB 223 ARR---KVAEALTIVVVMKRRREEEGAEKKMDLALLAADDGFSDEIYDF-----L 272

QY 295 TELFGGHETTASATSLITYGLYPHYLVQVREIKSKGLCKSNQDNKIDMETLEQLK 354

DB 273 VALLVAGYETTSTIMTLAVKFLTETPLALAOIKEE--HEKIRAMKSDSYLSLEWSDYKSM 330

QY 355 YICVVKETLRLNPPVPGGERVALKTFELNGYQIPKGMNVIYSICDTHDVADIFTNKEEF 414

DB 331 FTQCQVNETLRVANIIGVFRRAMTDEIKYKIPKWKVSSFRAYHLDPNHFKDARTF 390

QY 415 NPDRF---IVPHEDASRFSFIPEGGGLRSCVGEKFAKILKIFVELARHCDWQLNG 470

DB 391 NPMRWQSNSTVTGPSNV---FTPEGGGPRICPGYELARVALSVFLHRLVTGFSW----- 441

QY 471 PPTMKTSPTVYPVDNLPAFTYF 493

DB 442 VPAEQDKLVFFPTTRTQKRYPF 464

RESULT 7

T07859

Cytochrome P450 homolog - tomato

C:Species: Lycopersicon esculentum (tomato)

C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 16-Feb-2001

C/Accession: T07859

R:Bishop, G.J.; Harrison, K.; Jones, J.D.

Plant Cell 8, 959-969, 1996

A>Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes the

A:Reference number: Z16181; MUID:96266705

A/Accession: T07859

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-464 <BIS>

A:Cross-references: EMBL:U54770; NID:g1421740; PIDN:AAB17070.1; PID:g1421741

A/Experimental source: strain GCR758

C:Genetics:

A:Gene: dwarf

C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

F;273-436/Domain: cytochrome P450 homology <P45>

Query Match 16.9%; Score 442; DB 2; Length 464;

Best Local Similarity 27.0%; Pred. No. 9e-27;

Matches 125; Conservative 94; Mismatches 210; Indels 34; Gaps 10;

QY 18 LLLFLAALKLMDLYCVSSDRSCALPLPPGTMGFFFGETLQMVLRKRFLOMKRRKYGF 77

DB 14 LCIFCTALLRWNVQVKNQKN-----LPPGTMGWFLFGETTEFLKLGPSWMKNQARVGS 67

QY 78 IYKTHLFGRPYRVWGANVRRIILGHRVSVHWPASVRTILGAGCLSLNHDSSHQKQK 137

DB 68 FRKSHILGCPPTIVSMDSRLRYLVNEAKGLVPGYPOSMIDILGKCNIAVNGSAHKYMR 127

QY 138 KVIQAFSREALQCVLV-IAEVSSCLEQWLSCGERGLLYPEVKRLMIRIAMRILGCG 196

DB 128 GALSLISPTMIRDQLPRKIDEFMRSHLTNW---DNKVIDIQKTNKMAFLSSIKQIAGI 184

QY 197 EPGPAGGEGEQLVEAFEEEMTRNLFSLPIDVPFSGLYRGVKAR---NLHARIEENIR 252

DB 185 ESTSLA-----QEFMSEFFNLVLTSLPINLPNTNYHGFQARKIYNLRLILEE--R 237

QY 253 AKIRRLQATEPDGGCKDALQLLIEHSWGERGLDMQALKOSSSTELLFGGHETTASATSL 312

DB 238 RASKETQ-----HDMGLYLMNEATRFKLTDEMDILIT-ILSYETVSTTSMA 288

QY 313 IYIYGLYPHYLVQVREIKSKGLCKSNQDNKIDMETLEQLKYGCVIKETILRLNPPVPG 372

DB 289 VKYIHDHPKVLLELRKE--HMAIREKKRPEDPIDYNDYRSMRFTRAVILETSRLATVNG 346

QY 373 GFRVALKTFELNGYQIPKGMNVIYSICDTHDVADIFTNKEEFNPDRFIYPHEDASRFSF 432

DB 347 VLKRTTQDMEINGYIIPKGMRIYVYTRELNYDPRLYPDYPSFNPWRWMDKSLSHQNN--SF 404

QY 433 IPEGGGLRSCVGEKFAKILKIFVELARHCDWQLNGPPTMK 475

DB 405 LVFEGGTRQCPGKELGVAEISTFLHYFTKYRWEEIGGDKLMK 447

RESULT 8

H86185

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C/Accession: H86185

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A/Accession: H86185

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-490 <STO>

A:Cross-references: GB:AE005172; NID:g2388581; PIDN:AAB71462.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

[illegible]

RESULT 11

probable cytochrome P450 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: B84733
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: B84733
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <STO>
A:Cross-references: GB:AE002093; NID:g3831452; PIDN:AAC69934.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g32440
A:Map position: 2
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match	15.5%;	Score 406;	DB 2;	Length 489;
Best Local Similarity	27.8%;	Pred. No. 6.3e-24;		
Matches 128;	Conservative 80;	Mismatches 221;	Indels 32;	Gaps 12

[illegible]

```

Db      389  WKVLTFERNVHLDPEIYPDPKKEFDSRWEGYTPKAG--TFLEPGLGSHLCPGNDLAKLE 445
QY      452  LKIFTVELARHCDQLLN-GPPTMKTSPTVYPVDNLPART 491
Db      446  ISIFLHHEFLIKYVERSNPGCPVM-FLPHNRPKDNLARIT 485

```

RESULT 12

cytochrome P450 DWARF3 - maize
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Zea mays (maize)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: T02263
R:Winkler, R.G.; Helentjaris, T.
Plant Cell 7, 1307-1317, 1995
A:Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibber
A:Reference number: Z14648; MUID:96004534
A:Accession: T02263
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-519 <WIN>
A:Cross-references: EMBL:U32579; NID:g987266; PIDN:AAC49067.1; PID:g987267
A:Experimental source: strain B73
C:Genetics:
A:Gene: dwarf3
C:Function:
A:Description: involved in an early step in gibberellin biosynthesis
A:Pathway: gibberellin biosynthesis
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: oxidoreductase
F:325-488/Domain: cytochrome P450 homology <P45>

Query Match	13.7%;	Score 359;	DB 1;	Length 519;
Best Local Similarity	26.3%;	pred. No. 3.2e-20;		
Matches 123;	Conservative 89;	Mismatches 213;	Indels 42;	Gaps 15;

QY	44	LPRGTMGFPFEGCETLQMVLQRRK----	FLQMKRRKYG-FIKTHLFGRTVRYMAGDN	96
Dd	72	LPFGEMGWPLVGMMWAFLLRAFKSGRDAFIASFVRFRGRGVYRSEMFSSPTVLVTVAEG		131
QY	97	VRRILLGEHRLYSVHWMPASVRTILGACLSNLHDSHKQRKKIMQAFSR-EALQCYVLV		155
Dd	132	CQOYLMDDDAPVT-GWPKATVALVGRSFVAMPYDEHRIRIKTLTAAPINGFDALTGYLFF		190
QY	156	IAEEVSSCEBOWLSCGERGLTV-PEVKRLMRIAMRILLCEBPAGGEDE--QQLV		211
Dd	191	IDRTVTSILRAMADHG-GSVEFELELRMTFKIIVQIFL-----GGAQDATTRALE		240
QY	212	EAFEEMTRNLESLPIDVPFSGLYRG-VKARNLIHARIEENIRAKIRRLQATEPDDGCCDA		270
Dd	241	RSTELINYGMRAMAINLP-GEAYRGALRRARRLYA-VLQGVLDERRARAARAKVSGGVDM		298
QY	271	LOLLIEHSWERGERLDMQALKOSSTELLEFGHETTASAATSILTYLGlyPHVLOKVREI		330
Dd	299	MDRILEADERGRHLDDDEIIDLVMYMLNAGHESSGHITMATVFLOENPDMEARAKAEQ		358
QY	331	KSKGLCKSNQDNKLDMETLEQKYICVIKETRLRNPPVPGGFRVALKTFELNGYOIPK		390
Dd	359	EAIMRSIPSSQ-RGLTLRDFRKMEYLSQVIDETLRLVNISFVSFRQATRDFVNGYLTPK		417
QY	391	GWNVYISICTHDVADIFTNKKEFNDRFIVPHPEDASRFSTPFGGGLRSCVGEFAKI		450
Dd	418	GMKVQLWYRSYHMDPOVYPDPTEKEDSRWEHGSHPRAG--TFILAFLGARLCPSGNDLAKL		474
QY	451	LLKI FT-----VELARHCWQLNGBPTMKTSPTYVPDNLBARFT		491
Dd	475	EISVFLHHFLLGYKLAR-----TNPRCVRVRYLPHRPVDNCLAKIT		515
RESULT	13			

JC4240
lanosterol 14alpha-demethylase (EC 1.14.14.-) cytochrome P450 51 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 28-Jul-2000
C;Accession: JC4240; PC4069
R;Sloane, D.L.; So, O.Y.; Leung, R.; Scarafia, L.E.; Saldou, N.; Jarnagin, K.; Swinney,
Gene 161, 243-248, 1995
A;Title: Cloning and functional expression of the cDNA encoding rat lanosterol 14-alpha
A;Reference number: JC4240; MUID:95394364
A;Accession: JC4240
A;Molecule type: mRNA
A;Residues: 1-486 <SL0>
A;Cross-references: GB:U17697; NID:9699395; PIDN:AAA87074.1; PID:9699396
A;Accession: PC4069
A;Molecule type: protein
A;Residues: 43-60;206-213;254-267;369-387;420-428;433-443 <SL2>
A;Experimental source: liver
C;Comment: This enzyme is a cytochrome P-450 enzyme which catalyzes one of the essential
C;Superfamily: human cytochrome P450 CYP51; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; liver; metalloprotein; monooxygenase; oxidoreduct
F;291-454/Domain: cytochrome P450 homology <P45>
F;432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.0%; Score 339; DB 2; Length 486;
Best local Similarity 23.7%; Pred. No. 1.1e-18;
Matches 121; Conservative 97; Mismatches 230; Indels 62; Gaps 16;
QY 6 LLSAL--CTFVLP--LFLAALKIMDLVCVSSRDRSCALPLPPTMG-----PFPG 55
DB 9 LSTLLIACAFSLVYLFRLAV-----GHMVQLPAGAKSPRYISPIPLG 55
QY 56 ETLQWVLRKFLQMKRRKYGFIYKTHLFGRTVRVWGADNVRRLIGEHRLVSVH--WP 113
DB 56 HAIAFGKSPIEFLENAYEKYGVFSFTVWGKFTTYLLGSDAALLFNSKNEDLNAEVYG 115
QY 114 ASVRTILGAGCISNLHDSHKQKKVIMQAFSREALQCVLVIAEVSLSCEQWLSGGER 173
DB 116 RLTPVFGKGVAYDVPNAVFLQKILKSLGSLNIAHFQYVSIIIEKAKEYFKSWGESGER 175
QY 174 GLVYPEVKRLMFRIRMLLGCPEPGAGGEDEQQLVEAFEMTRNL--FS-----LP 225
DB 176 N--VFELSELILLTASHCL-----HGKEIRSQLEKVAQLYADLDGGFSHAAMLLP 225
QY 226 IDVPFSGLYRGVKA---RNLIHARIEENIRAKIRRLQATEPDGCKDALOLLIEHSWER 281
DB 226 GWLPPLPSFRRDRRAHREIKNIFFKAIOK-----RRL-SKEP--AEDILQTLDDSTYKD 275
QY 282 GERLDMQALKQSTELLFGGHETTASATSLITYLGLYPHYLVQKVEEIKSKGLCKSNQ 341
DB 276 GRPLTDEIAGMLIGLLLAGQHTSSTTSAMWGFLLARDKPLQDKCYLEOKT--VCGEDL 332
QY 342 DNKLDMETLEQLKYIGCVIKETRLRNPVPGGFRVALKTFELNGYQIPKGMNVISICDT 401
DB 333 P-PLTYEQLKDLNLDRCIKETRLRNPIMTMMAKTPQTVAGYTIIPGHOCVSPVYN 391
QY 402 HDVADIFTNKEENPDREIVPHPEDASRFSFIPEGGLRSCVGEKFAKILKIFTVELAR 461
DB 392 QRLKDSWVERLDENPDRLQDNPASGEKFAVYVFGAGRHRICGENFAVYQIKTIWSTMLR 451
QY 462 HCDWQLNGP-PTMKTSPTVYPVDNLPAKF 490
DB 452 LYEDLINGYFSPVNYTMTHTPENPVIRY 481

RESULT 14
JC4758
lanosterol 14alpha-demethylase (EC 1.14.14.-) cytochrome P450 51 - rat
N;Alternate names: pRt-9 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 28-Jul-2000
C;Accession: JC4758; PC4170; JC5888; JC2334
R;Aoyama, Y.; Noshiro, M.; Gotoh, O.; Imaoka, S.; Funae, Y.; Kurosawa, N.; Horiuchi, T.;

J. Biochem. 119, 926-933, 1996
A;Title: Sterol 14-demethylase P450 (P45014DM) is one of the most ancient and conserv
A;Reference number: JC4758; MUID:96389999
A;Accession: JC4758
A;Molecule type: mRNA
A;Residues: 1-503 <AOY1>
A;Cross-references: DDBJ:D55681; DDBJ:D29962; NID:q1020094; PIDN:BAA09529.1; PID:9870
A;Accession: PC4170
A;Molecule type: protein
A;Residues: 52-109;305-319;324-334;467-482;488-499 <AOY2>
A;Experimental source: liver
R;Noshiro, M.; Aoyama, Y.; Kawamoto, T.; Gotoh, O.; Horiuchi, T.; Yoshida, Y.
J. Biochem. 122, 1114-1121, 1997
A;Title: Structural and evolutionary studies on sterol 14-demethylase P450 (CYP51), t
A;Reference number: JC5888; MUID:98158318
A;Accession: JC5888
A;Molecule type: DNA
A;Residues: 1-503 <NOS>
A;Cross-references: DDBJ:AB004087
R;Aoyama, Y.; Funae, Y.; Noshiro, M.; Horiuchi, T.; Yoshida, Y.
Biochem. Biophys. Res. Commun. 201, 1320-1326, 1994
A;Title: Occurrence of a P450 showing high homology to yeast lanosterol 14-demethylas
A;Reference number: JC2334; MUID:94296405
A;Accession: JC2334
A;Molecule type: mRNA
A;Residues: 74-503 <AOY>
A;Cross-references: DDBJ:D29962
A;Experimental source: liver
C;Genetics:
A;Gene: CYP51
A;Introns: 59/3; 91/3; 150/3; 193/1; 251/2; 291/2; 356/3; 388/3; 445/1
C;Function:
A;Description: catalyzes the removal of the 14-methyl group of 14-methylsterols
C;Superfamily: human cytochrome P450 CYP51; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; liver; metalloprotein; monooxygenase; oxidore
F;308-471/Domain: cytochrome P450 homology <P45>
F;449/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.0%; Score 339; DB 2; Length 503;
Best local Similarity 23.7%; Pred. No. 1.1e-18;
Matches 121; Conservative 97; Mismatches 230; Indels 62; Gaps 16;

QY 6 LLSAL--CTFVLP--LFLAALKIMDLVCVSSRDRSCALPLPPTMG-----PFPG 55
DB 26 LSTLLIACAFSLVYLFRLAV-----GHMVQLPAGAKSPRYISPIPLG 72
QY 56 ETLQWVLRKFLQMKRRKYGFIYKTHLFGRTVRVWGADNVRRLIGEHRLVSVH--WP 113
DB 73 HAIAFGKSPIEFLENAYEKYGVFSFTVWGKFTTYLLGSDAALLFNSKNEDLNAEVYG 132
QY 114 ASVRTILGAGCISNLHDSHKQKKVIMQAFSREALQCVLVIAEVSLSCEQWLSGGER 173
DB 133 RLTPVFGKGVAYDVPNAVFLQKILKSLGSLNIAHFQYVSIIIEKAKEYFKSWGESGER 192
QY 174 GLVYPEVKRLMFRIRMLLGCPEPGAGGEDEQQLVEAFEMTRNL--FS-----LP 225
DB 193 N--VFELSELILLTASHCL-----HGKEIRSQLEKVAQLYADLDGGFSHAAMLLP 242
QY 226 IDVPFSGLYRGVKA---RNLIHARIEENIRAKIRRLQATEPDGCKDALOLLIEHSWER 281
DB 243 GWLPPLPSFRRDRRAHREIKNIFFKAIOK-----RRL-SKEP--AEDILQTLDDSTYKD 292
QY 282 GERLDMQALKQSTELLFGGHETTASATSLITYLGLYPHYLVQKVEEIKSKGLCKSNQ 341
DB 293 GRPLTDEIAGMLIGLLLAGQHTSSTTSAMWGFLLARDKPLQDKCYLEOKT--VCGEDL 349
QY 342 DNKLDMETLEQLKYIGCVIKETRLRNPVPGGFRVALKTFELNGYQIPKGMNVISICDT 401
DB 350 P-PLTYEQLKDLNLDRCIKETRLRNPIMTMMAKTPQTVAGYTIIPGHOCVSPVYN 408
QY 402 HDVADIFTNKEENPDREIVPHPEDASRFSFIPEGGLRSCVGEKFAKILKIFTVELAR 461
DB 402 HDVADIFTNKEENPDREIVPHPEDASRFSFIPEGGLRSCVGEKFAKILKIFTVELAR 461

Db 409 QRLKDSWVERLDENPDRLQDNPASGEKFAVYPCAGRHRCIGENFAVYQIKTIWSTMLR 468
QY 462 HCDWQLNGP-PTMKTSPYVPVNDLPARF 490
Db 469 LYEFDLINGYFPSVNTTMIHTPENPVIRY 498

Search completed: November 6, 2001, 13:35:31
Job time: 599 sec

RESULT 15
T04591
ferulate-5-hydroxylase (EC 1.-.-.-) - Arabidopsis thaliana
N;Alternate names: cytochrome P450-dependent monooxygenase; protein F23E13.110
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 28-Jul-2000
C;Accession: T04591
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.
submitted to the Protein Sequence Database, March 1998
A;Reference number: Z15378
A;Accession: T04591
A;Molecule type: DNA
A;Residues: 1-520 <BEV>
A;Cross-references: EMBL:AL022141
A;Experimental source: cultivar Columbia; BAC clone F23E13
C;Genetics:
A;Map position: 4
A;Introns: 171/1; 314/3
A;Note: F23E13.110
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: heme; iron; metalloprotein; oxidoreductase
F;316-480/Domain: cytochrome P450 homology <P45>
F;458/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 12.9%; Score 337; DB 2; Length 520;
Best Local Similarity 25.8%; Pred. No. 1.7e-18;
Matches 133; Conservative 90; Mismatches 194; Indels 98; Gaps 23;
QY 15 VLPLLFLAALKLWLDYCVSSRDRSCALPLPGTMGPFPGETLQWLVQRKFLQMKRRK 74
Db 22 VVSLFIFIS-----FTRRRRP--PYPPGRGWPIIGNMLMDQLTHRGLANLAKK 70
QY 75 YGFIYKTHL-FGRPTVRVNGADNVRRIILGHRIVSVHWPASVRTILGAGCLSNL-HDSS 132
Db 71 YGGL--CHLRMGFLHMYAVSSPEVARQVLQVDSVFSNRPATI-----AISYLYTDRA 121
QY 133 -----HKQRKVIMQAFSREALQCYLV-----IAEYSSCLEQWLSGGERGL 175
Db 122 DMAFAHYGPFWRQMRKYCVMKVFSRKRAESWASYRDEVDMKVRSVSCNKGKPINVGEO-- 179
QY 176 LVYPEVKRLMFIAMRIILGCEPAGGGEQOLVEAFEEETR--NLFSPIDIVPFSGL 233
Db 180 -IFALTRNITYRAAFG--SACEKG-----QDEFIRILOEFSKLFGAFNVADEFIPYFGW 229
QY 234 Y-----RGVKARNLIHARIEENI--RAKIRILOATEPDG-----CKDALQLLIEHS 278
Db 230 IDPOGINKRLVKARNLDLGFIDDIIDEHMKKENQNAVDDGDVDTDMVDDLLAFYSEEA 289
QY 279 WERGERLDMQ-----ALKQSTELLFGHETTASATSLITYLGLYPHYLQKVREI 330
Db 290 KLVSETADLQNSIKLTRDNKAIIMDMVFGTETVASAIEWALTELLRSPEDLKRVOQEL 349
QY 331 -KSKGLCKSNQDNKLDMETLEQLKYGVCVIRKTLRLNPPVPGGFRVALKTFELNGYQIP 389
Db 350 AEVYGL-----DRVVEESDIEKLTLYLCKTLEKTLRMHPPIPLLHETAEDTSIDGFEIP 403
QY 390 KGMNVIYSICDTHVDADIFTNKEEENPDREIVPHED--ASRESFIPGGGIRSCVGEF 447
Db 404 KKSRRVMINAFAIGRDPTSWTDPTRFSPRLFEPGVDFKGSNFEFIPFGSGRRSCPGMQ- 462
QY 448 AKLLKIFTVELA---RHC-DWQLNGPPTMKT 477
Db 463 ----LGLYALDLAVAHILHCTWKL--PDGMKPS 490

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:38:27 ; Search time 115.92 Seconds
(without alignments)
567.250 Million cell updates/sec

Title: US-09-668-482-32
Perfect score: 2612
Sequence: 1 MGLPALLASALCTFVLPLL.....PTVYPVDNLPAKFTYFGDI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2609	99.9	497	11 Q9R1F4	Q9R1F4 mus musculu
2	2057.5	78.8	492	13 Q9PUB4	Q9PUB4 gallus gall
3	1758.5	67.3	492	13 Q93323	Q93323 xenopus lae
4	1048.5	40.1	512	4 Q9NR63	Q9NR63 homo sapien
5	1036	39.7	525	4 Q9NP41	Q9NP41 homo sapien
6	668.5	25.6	444	2 Q59990	Q59990 synechocyst
7	569	21.8	477	10 Q9LVY7	Q9LVY7 arabidopsis
8	538	20.6	107	13 Q9PUG2	Q9PUG2 gallus gall
9	517.5	19.8	463	10 Q9FH76	Q9FH76 arabidopsis
10	511.5	19.6	485	10 Q9SJH2	Q9SJH2 arabidopsis
11	501.5	19.2	482	10 Q81077	Q81077 arabidopsis
12	494.5	18.9	457	10 Q65624	Q65624 arabidopsis
13	461	17.6	513	10 Q64989	Q64989 arabidopsis
14	461	17.6	513	10 Q9SCQ9	Q9SCQ9 arabidopsis
15	455.5	17.4	465	10 Q9FMA5	Q9FMA5 arabidopsis
16	432	16.5	443	10 Q9LJK2	Q9LJK2 arabidopsis
17	418.5	16.0	465	10 Q9LH81	Q9LH81 arabidopsis
18	418	16.0	457	10 Q9M066	Q9M066 arabidopsis
19	418	16.0	474	10 Q9LKH7	Q9LKH7 phaseolus a

20	418	16.0	524	10 Q23242	Q23242 arabidopsis
21	407.5	15.6	464	10 Q9LIC5	Q9LIC5 arabidopsis
22	406	15.5	489	10 Q9ZV72	Q9ZV72 arabidopsis
23	404	15.5	478	10 Q9LN73	Q9LN73 arabidopsis
24	400.5	15.3	496	10 Q9FQY4	Q9FQY4 cucurbita m
25	380	14.5	475	2 Q9X7G9	Q9X7G9 myxococcus
26	374	14.3	504	10 Q9SNG3	Q9SNG3 oryza sativ
27	358.5	13.7	511	10 Q9SWR1	Q9SWR1 liguidambar
28	351	13.4	486	11 Q9JTY3	Q9JTY3 mus musculu
29	351	13.4	503	11 Q9JIP8	Q9JIP8 mus musculu
30	350.5	13.4	513	10 Q9FVB8	Q9FVB8 brassica na
31	342	13.1	520	10 Q9FVC0	Q9FVC0 brassica na
32	341	13.1	520	10 Q9FVB9	Q9FVB9 brassica na
33	340	13.0	318	10 Q9LVY3	Q9LVY3 arabidopsis
34	336.5	12.9	735	10 Q9XFM2	Q9XFM2 lycopersico
35	336	12.9	521	10 Q9XFM2	Q9XFM2 lycopersico
36	333	12.7	518	10 Q9FI38	Q9FI38 arabidopsis
37	329.5	12.6	512	10 Q9FX29	Q9FX29 arabidopsis
38	327	12.5	498	5 Q16806	Q16806 drosophila
39	318.5	12.2	474	10 Q9SHY7	Q9SHY7 arabidopsis
40	318.5	12.2	491	5 Q9NGK3	Q9NGK3 tribolium c
41	318.5	12.2	513	10 Q9SBP8	Q9SBP8 populus trt
42	318	12.2	497	5 Q76478	Q76478 diptoptera
43	317.5	12.2	459	2 Q59163	Q59163 anabaena va
44	315.5	12.1	495	10 Q9TOK2	Q9TOK2 arabidopsis
45	314	12.0	464	10 Q04949	Q04949 arabidopsis

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	497 AA.
Q9R1F4	Q9R1F4			
AC	Q9R1F4			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	CYTCHROME P450 RETINOIC ACID METABOLIZING ENZYME P450RA.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=TOOTH;			
RX	PubMed-11063033;			
RA	Paine C.T., Paine M.L., Snead M.L.;			
RT	"Identification of tuftelin- and amelogenin-interacting proteins using			
RT	the yeast two-hybrid system."			
RL	Connect. Tissue Res. 38:257-267(1998).			
CC	- - CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +			
CC	OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).			
CC	- - SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY			
CC	SIMILARITY).			
CC	- - SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
DR	EMBL; AF115769; AAD17217.1; -.			
DR	InterPro; IPR001128; -.			
DR	Pfam; PF00067; P450; 1.			
DR	PRINTS; PR00385; P450.			
DR	PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.			
KW	electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;			
KW	Monooxygenase; Oxidoreductase.			
SQ	SEQUENCE 497 AA; 56191 MW; 67CBI67A997842C1 CRC64;			

Query Match 99.9%; Score 2609; DB 11; Length 497;
Best Local Similarity 99.8%; Pred. No. 1.3e-205;
Matches 496; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1 MGLPALLASALCTFVLPLLLFLAALKMLDIXCVSSRDRSCALPLPPTGTMGFPFGFTLQM 60
DB	1 MGLPALLATATCTFVLPLLLFLAALKMLDIXCVSSRDRSCALPLPPTGTMGFPFGFTLQM 60

QY	61	VLORRKFLÖM	KRRKYYGFIYKTHLFGRPTVYMGADNVRRILLGEHRLVSVMH	PASVRTIL	120		
Db	61	VLORRKFLÖM	KRRKYYGFIYKTHLFGRPTVYMGADNVRRILLGEHRLVSVMH	PASVRTIL	120		
QY	121	GAGCLSNLHDS	SHKÖRKVKVIMQAFSREALÖCYVLVIAEEVSSCLEQWLS	CGEGLLVPE	180		
Db	121	GAGCLSNLHDS	SHKÖRKVKVIMQAFSREALÖCYVLVIAEEVSSCLEQWLS	CGEGLLVPE	180		
QY	181	VKRIMFRIAMR	ILLGCEP	PAGGEGEÖQVLEAFEEMTRNLFSLPIDVPSGLYRGVKAR	240		
Db	181	VKRIMFRIAMR	ILLGCEP	PAGGEGEÖQVLEAFEEMTRNLFSLPIDVPSGLYRGVKAR	240		
QY	241	NLIHARIEENIRAKIR	RLQATEP	DGCKDALÖLLIEH	SWERGERLDMQALKÖSSTELLFG	300	
Db	241	NLIHARIEENIRAKIR	RLQATEP	DGCKDALÖLLIEH	SWERGERLDMQALKÖSSTELLFG	300	
QY	301	GHETTASAATSLIT	YLGYPHVLÖK	VREIEIKSKGLLCKSNÖDNK	LDMETLEÖLKYIGCVI	360	
Db	301	GHETTASAATSLIT	YLGYPHVLÖK	VREIEIKSKGLLCKSNÖDNK	LDMETLEÖLKYIGCVI	360	
QY	361	KETLRLNRPV	PGGF	RVALKTFELNGYÖIPK	GNVIYISICDTHVADIFTNKEEFN	PDREI	420
Db	361	KETLRLNRPV	PGGF	RVALKTFELNGYÖIPK	GNVIYISICDTHVADIFTNKEEFN	PDREI	420
QY	421	VPHPEDASRFS	FIPEGGGLRSCV	GKEFAKILKIFTVELARHCDMÖLLNG	PPTMKTSP	TV	480
Db	421	VPHPEDASRFS	FIPEGGGLRSCV	GKEFAKILKIFTVELARHCDMÖLLNG	PPTMKTSP	TV	480
QY	481	YPVDNLPARFTY	FOGDI	497			
Db	481	YPVDNLPARFTY	FOGDI	497			

```

RESULT      2
O9PUB4
ID O9PUB4      PRELIMINARY;      PRT;      492 AA.
AC O9PUB4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RAI) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBL_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Swindell E.C., Thaller C., Sockanathan S., Petkovich M., Jessell T.M.,
RA Eichele G.;
RT "Complementary Domains of Retinoic Acid Production and Degradation in
RT the Early Chick Embryo.";
RL Dev. Biol. 0:0-0(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF199462; AAF09250.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 492 AA; 55264 MW; 7F28B72E75C232FB CRC64;

```

Query Match	Score	DB	Length
Best Local Similarity	2057.5;	13;	492;
	80.1%;	Pred. No. 2.1e-160;	

Matches	398;	Conservative	40;	Mismatches	54;	Indels	5;	Gaps	3;
---------	------	--------------	-----	------------	-----	--------	----	------	----

QY	1	MLPALLASALCTFVLPLLLFLAALKLMDLYCVSSRDRSCALPLPRGTMGPFPEFTLQ	60
Db	1	MGFSALVASALCTFLPLPLFLAAYRLMDLYCASCGRDPSCLPLPRGTMLPFFGETLQ	60
QY	61	VLQRRKFLQMKRRKYGFYKTHLFGRPYRVMGADNVRILLGEHRLVSVHWPASVRTIL	120
Db	61	VLQRRKFLQMKRRKYGFYKTHLFGRPYRVMGAEENVRHILLGEHRLVSVQWPASVRTIL	120
QY	121	GAGCLSNLHDSHKQKRVIMQAFSREALQCYVLVAEEVSSCLEQWLSCGERGLTYPE	180
Db	121	GSGCLSNLHNGQHKHRKKVIMQAFSRDALQHYVPVLOEEVSACLAQWLGAGP-CLTYPE	179
QY	181	VKRLMFIAMRILLGCEPAGGGEDEQOLVEAFEEEMTRNLSLPIDVPFSGLYRGYKAR	240
Db	180	VKRLMFIAMRILLGFQPRQ-SPDGEQOLVEAFEEEMIRNLSLPIDVPFSGLYRGELAR	238
QY	241	NLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIHSWGERGLDMQALKOSSTELLFG	300
Db	239	NIHAKIEENIRAKMAR--KEPEGYKDALQLLMEHTQNGEOQLNMQELKESATELLFG	295
QY	301	GHETTSASATSLITYLGLYPHVLOKVREEIKSKGLCKSNQDNKLDMETLEQLKYICVY	360
Db	296	GHETTSASATSLIAFLGLHHDVLOKVREKELQKGLLSGPNQEKQLNMEFLEQLKYTCVY	355
QY	361	KETLRLNPVPVPGGFRVALKTFELNGYQIPRGWNVYISICDTHDVADIFTNKEEFNDRFI	420
Db	356	KETLRLSPVPVPGGFRIALKLTLELNGYQIPRGWNVYISICDTHDVADLFTDKDEFNDRFM	415
QY	421	VPHPEDASRFSFIPEGGGLRSCVCGKEFAKILLKIFTVELARHCDWQLNGPPTMKTSPV	480
Db	416	SPSEEDSSRFSFIPEGGGLRSCVCGKEFAKILLKIFTVELARSCDWQLNGPPTMKTGPV	475
QY	481	YPVDNLPARFTYFQGDY 497	
Db	476	YPVDNLPARFTYFSGOI 492	

```

RESULT      3
O93323
ID   O93323          PRELIMINARY;      PRT;      492 AA.
AC   O93323;
DT   01-NOV-1998 (TREMBLrel. 08, Created)
DT   01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE   CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE   CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chen J.L., Grunz H., Panitz F., Pieler T., Hollemann T.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC   -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
EMBL AF057566; AAC25158.1; -.
DR   InterPro; IPR001128; -.
DR   Pfam; PF00067; P450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE 492 AA; 55459 MW; D1D4BB7651BF2D3E CRC64;

```

```
Query Match          67.3%; Score 1758.5; DB 13; Length 492;  
Best Local Similarity 68.1%; Pred. No. 6.9e-136;  
Matches 339; Conservative 63; Mismatches 89; Indels 7; Gaps 4;  
  
QY      1 MGPPALLASALCTFVLPLLLFLAALKLMDLYCVSSRRDSCALPUPPGTMGFPEGTIQQM 60  
| | | | | | | | | | | | | | : | : | : | | | | | | | | | |
```

```
Db      1 MDLYTLTSLALCTLAPLLPLLTAAKLMEVYCLRRKDAACANLPDPTMGLPFFGETIOM   60
Qy      61 VLQRKRFLQMRRKYGFITYKTHTLFGRPTVRVMGADNVRILLGEHRLVSVMHPASVRTIL  120
        |||||:::||::|| | | | | | | | | | | : |:| ::||:| ::||| | | | | | 
Db      61 VLRRRRFLQVKRSQYGRITKTHLFSGPTVRVTGAENVROQLMGHEKLVSVHWPA SVRTIL  120
Qy      121 GAGCLSNLDSSHKORKKVIMQAFAFSREALQCYYVLVIAEVEVSSCLEOWLSGGERGLVYPE  180
        |||||||::| | | | | | | | | | | | : | | | : | | | | | 
Db      121 GAGCLSNLDHNHHKKYTCKVIAQAFSREALANYVPOMEEEVRCVNMMLQSGP-CVLVYP A  179
Qy      181 VKRLMFRIAMRILLGCEPPAGGGEGDEOOLVEAFEEMTRNLFSLPIDVPFSGLYRGVKAR  240
        ::|||::| | | | | | | | | | | | : | : | | | | | | | | | | | | | | 
Db      180 IKRMFMFIAMRILLGC DP-QRMDREQEETLLAEFEEMS RNLFSLPIDVPFGSLYRG LRAR  238
Qy      241 NLIHARIENIRAKIRRLQATEPDGGCKDALQLLIEHSWARGERDLMOALKOSTELLFG    300
        |||||::| | | | | | | | | | | | : | | | | | | | | | | | | | | | 
Db      239 NLIIHAQIEENTKEKLOR---EPDECHKALQOLLIDYSRRNGEPIINQALKE SATELLFG  294
Qy      301 GHETTASAATSLTYLGlyPHYLQKVREETSKGLL-CKSNODNKLDMETLEQLKIYIGCV    359
        || ||||| | | | | | | | | | | | | : | : | | | | | | | | | | | | 
Db      295 GHGTASAATSISFLALHKVDLEKVRKETEQGLLSTKP EKKELSTEVLQQLKYTS CV     354
Qy      360 IKETLRLNPPVGGEFRVALKTFELNGYQIPKGWNVIYSICDTHADVADIFTNKEEFNPDR F   419
        |||||||::| | | | | | | | | | | | | | | | | | | | | | | | | | | 
Db      355 IKETLRISPVAAGEFRVALKTFVLNGYQIPKGWNVIYSI ADTHGEADLEFPDTDKFN PDR F   414
Qy      420 IVPHPEDASRFSEIFPGGGGLRSCVCGKEFAKILIKFTVELARHCWDQLNGPPTMTKS PT    479
        : | | : | | | | | | | | | | | | | | | | | | | | | | | | | | 
Db      415 LTPLPRDSSRFGEIFPGGGVCIGCIGKEFAKILIKVFVELCRNC DWMELLNGSPAMTT S PI   474
Qy      480 VYPVNDLPA RF T YFOGDI 497
        : |||||::| | | | | | | | | | | | | | | | | | | | | | | | | | 
Db      475 ICPVDNLPAKF KPFSSSI 492

RESULT          4
Q9NR63 PRELIMINARY; PRT; 512 AA.
ID Q9NR63
Ac Q9NR63;
Dt 01-OCT-2000 (TREMBLrel. 15, Created)
Dt 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
Dt 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
De CYTOCHROME P450 RETINOID METABOLIZING PROTEIN P450RAI-2.
GN CYP26B1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20300913; Pubmed=10823918;
RA White J.A., Ramshaw H., Talmi M., Stangle W., Zhang A., Everingham S.,
RA Creighton S., Tam S.-P., Jones G., Petkovich M.;
RT "Identification of the human cytochrome P450, P450RAI-2, which is
RT predominantly expressed in the adult cerebellum and is responsible for
RT all-trans-retinoic acid metabolism.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:6403-6408(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC CC SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
Cc -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
Cc EMBL; AF252297; AAF76003.1; -.
Dr InterPro; IPR001128; -.
Dr Pfam; PF00067; p450; 2.
Dr PRINTS; PR00385; P450.
Dr PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 512 AA; 57512 MW; A06D1D9944E6726F CRC64;
```

Query Match 40.1%; Score 1048.5; DB 4; Length 512;

	Best Local Similarity	44.48;	Pred. NO. 1.2e-77;						
	Matches	220;	Conservative	88;	Mismatches	164;	Indels	23;	Gaps
									9;
QY	7	LASALCTFVLP----	LLFLALKLMDLYCVSSSRDRSCALPLPGTMGFPFFGETLQMWL	62					
Db	8	LVSALATLACLAVSVTL	LAVSQDMLRWATROKSKCLPIPKSGMGLIGETGHWLL	67					
QY	63	QRRKFLQMKRRKRYGY	IKTTHLEGRPTVRVMGADNVKRILIGEHRLVSVHWPASVRIILGA	122					
Db	68	QSGGFQSSRREKRYGN	VEKTHLLGRPLIRVTGAENVRKILMGHHLVSTEMPFRSTMLGP	127					
QY	123	GCLSNLHDSHKORKK	VIMQAFSREALQCYLVIAEEVSSCLEQWLSCGGERGLVYPEVK	182					
Db	128	NTVSNISIGDIHRNKR	KVFSKIFSHAELESYLPKIQLVLODTRLAWSSHE-AINYYQEAQ	186					
QY	183	RLMERIAMRIILGCE	-PGPAGGGEDEQOLVEAFEEMTRNLESLPIDVPFSGLYRGYKARN	241					
Db	187	KLTFRMAIRVLLGFS	IP-----EEDLGHLEFVYQOFVDNVFSLPVDLPFGSGYRRGIGIARQ	241					
QY	242	LIHARIEENIRAKIR	RLQATEPDDGGCKDALQLLIEHSWERGEKRLDMQALKOSSTELLEFGG	301					
Db	242	ILQGLEKAIKEKIQ	CTQCKD----YLDALDLLIESSKEHGEKEMTQELKDGTLLELFAA	297					
QY	302	HETTASAATSLITYL	GLYPHVLQKVREELTKSKGLL---CKSNQDNKLDMETLEQLKYG	357					
Db	298	YATTASASTSLIMQ	LKHPTVLEKLRDELRAHGILHSGGCP--EGTLRLDTLSGLRIYD	355					
QY	358	CVIKETLRLNPYPG	GFRVALKTFEELNGIPIKGNVITYSICDTHDVADIFTNKEEENPD	417					
Db	356	CVIKEVMRLFTPI	SGGYRTVLQTFELDFQIPKGSVMYSIRPDHTAPVEKDVNVEDPD	415					
QY	418	RFIVPHPEDA-SRFS	FIPFGGLRSCVGRKEFAKILLKIFTVELARHCDWQLNGP-PTMK	475					
Db	416	RFQQRSEDKDGRF	HYLPFGGAVRTCLGNLAKLELVIAVELASTSRFELATRTFPRIT	475					
QY	476	TSPTYVPVDNLPA	RF 490						
Db	476	LVPVLHPVDGLSV	KF 490						

```

RESULT      5
09NP41
ID      09NP41      PRELIMINARY;      PRT;      525 AA.
AC      09NP41;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT      01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE      WUGSC:H_NH0493L16.1 PROTEIN.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99063792; PubMed=9847074;
RA      Sulston J.E., Waterston R.;
RT      "Toward a complete human genome sequence.";
RL      Genome Res. 8:1097-1108(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Bourne S., Bauer C., Pape K., Jones T.;
RT      "The sequence of Homo sapiens BAC clone RP11-493L16.";
RL      Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RA      Waterston R.;
RL      Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC      -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC      OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC      SIMILARITY).
CC      -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR      EMBL; AC007002; AAF65576.1; -.

```

DR EMBL; AC007002; AAF65576.1; ..

40.18; Score 1048.5; DB 4; Length 512;

DR InterPro; IPR001128; -
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Mitochondrion;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 525 AA; 59124 MW; 71D47B6752A60315 CRC64;

Query Match 39.7%; Score 1036; DB 4; Length 525;
Best Local Similarity 43.3%; Pred. No. 1.3e-76;
Matches 220; Conservative 89; Mismatches 163; Indels 36; Gaps 10;

QY 7 LASALCTFVLP-----LFLAALKLMDLYCVSSRDSCALPLPPGTMGPPFEGTLMQVL 62
DB 8 LVSALATLAACLVSVTLTLLAVSQQLMWLRMAATRDKSCKLPKSGMGFPILGETGHML 67
QY 63 QR-----RKFLQMKRRKRYGTYKTHLFGRTVRVMGADNVRILLGEHRLVS 109
DB 68 QKCTLRRTVWLPQSGSGFOSSRREKYNFKTHLGRPLIRVTGAENVKILMGEHLVS 127
QY 110 VHWPAVFTILGAGCLSNLHDSHKORKVIMQAFSREALQCYVLVAEVSCLQWLS 169
DB 128 TEMPRSTRMLGPNVTVNSIGDIHRNKRKYSKIFSHEALSYLPKIQVLQVITDLRAWSS 187
QY 170 CGERGLVYPEVKRLMFRIMRILGCE-PGPAGGGEDEQQLVEAFEMTRNLSLPIDV 228
DB 188 HPE-AINVYQEAQKLTFRMAIRVLGSP-----EEDLGHLEFYVQOFVDNVSFLPVDL 241
QY 229 PFGSLYRGVAKARNLIHARIEENIRAKIRLQATEPDDGCKDALQLLIEHSWGERLDMQ 288
DB 242 PFGYRKGIOARQIIOKGLKAIKREKIOCTQGD----YLDALDLIESSKEHKEMTMQ 297
QY 289 ALKOSTELLEGGHETTASAATSLITYLGLYPHYLQKVREIKSGLL-----CKSNQDNK 344
DB 298 ELKQGTLELIFAAYATASASTSLIMQLKHPTVLEKLRDELRAHGILHSGGCP--EGT 355
QY 345 LDMETLEQLKYIGCVIKETLRNPNPVGGRVALKTFELNGYQIPKGNVITYSICDTHDV 404
DB 356 LRDLTLGLRLYLDVKEVMRLFTPISGGYRTVLQTFELDGFQIPKGSVMYSIRDTHDT 415
QY 405 ADIFTNKEENPDRTFVHPEDA-SRFSIFPFGGLRSCVGEKFAKILKIFTVELARHC 463
DB 416 APYFKDVNVEDPDRFSQARSEDKGRFHYLPFGGVRICLGHKLAKLEKLAVELASTS 475
QY 464 DWQLNGP-PTMKTSPVTPVDNLPAF 490
DB 476 RELATRTFPRITLVPVLPVLDGLSVK 503

RESULT 6
Q59990 PRELIMINARY; PRT; 444 AA.
AC Q59990;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE CYTOCHROME P450 120.
OS CYP120 OR CYP OR SLR0574.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA Tabata S.;
RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugiura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; D64003; BAI0496.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Hypothetical protein; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 444 AA; 50578 MW; 8F62A9EED3B54BDC CRC64;

Query Match 25.6%; Score 668.5; DB 2; Length 444;
Best Local Similarity 34.1%; Pred. No. 1.3e-46;
Matches 154; Conservative 87; Mismatches 181; Indels 29; Gaps 8;

QY 41 ALPLPGTMGPPFEGTLMQVLRKFLQMKRRKRYGTYKTHLFGRTVRVMGADNVRRI 100
DB 10 SLPIPGDFGLPWLGETLNF-LNDGFGKRRQOQFGPIFKTRLGKNVIFISGALANRL 68
QY 101 LLGHRIVSVHWPAVFTILGAGCLSNLHDSHKORKVIMQAFSREALQCYVLVAEEV 160
DB 69 FTKEQETFOATWPLSTRILLGNALATQMGIEIHSRKKILYQAFILPRLDSYLPKMDGIV 128
QY 161 SSCLQWLSGGERGLVYPEVKRLMFRIMRILGCEPGPAGGGEDEQQLVEAFEMTRN 220
DB 129 QGYLEQWKANE--VIWYPQLRRMTFDVAATLFMEKV-----SQNPQLPWFETIYG 180
QY 221 LFSLPIDVPSGLYRGVAKARNLIHARIEENIRAKIRLQATEPDDGCKDALQLLIEHSWE 280
DB 181 LFSLPIDVPSGLYRGVAKARNLIHARIEENIRAKIRLQATEPDDGCKDALQLLIEHSWE 280
QY 281 RGERLDMQALKOSTELLEGGHETTASAATSLITYLGLYPHYLQKVREIKSGLLCKSN 340
DB 235 NNQPLSLPELKDQILLFAGHEFTLSALSSFCLLGQHSIDREVRQEQN-----KLQ 288
QY 341 QDNKDMETLEQLKYIGCVIKETLRNPNPVGGRVALKTFELNGYQIPKGNVITYSICD 400
DB 289 LSGELTAETLKMPYLDQVLEVRLRIPVGGFRELIDQCFQGFHPKGLVSYQISQ 348
QY 401 THDVADIFTNKEENPDRTFVHPEDASR----FSIFPFGGLRSCVGEKFAKILKIFT 456
DB 349 THADPDLPDPEKDFPERFT---PDGSATHNPFAHVHFGGLRECLGKFAKILKIFA 405
QY 457 VELARHCDWQLNGPP-TMKTSPVTPVDNL 486
DB 406 TRLIQQFDWTLPLPGQNLVTPSPRPKDNL 436

RESULT 7
Q9LVY7 PRELIMINARY; PRT; 477 AA.
AC Q9LVY7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYTOCHROME P450-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20181125; Pubmed-10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:31-63(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB018112; BAA96885.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 477 AA; 54850 MW; 304B4B2C4970E405 CRC64;

Query Match 21.8%; Score 569; DB 10; Length 477;
Best Local Similarity 31.6%; Pred. No. 2.1e-38;
Matches 154; Conservative 82; Mismatches 215; Indels 36; Gaps 14;

OY 18 LLLFLAALKLMDLYCVSSRRSCALP-LPPGTMGFPFGETIQMLQRR----KFLQMK 71
Db 7 ILLFLSILLSLLLLRKHLSHFSYPNLPBGNTGLPLIGESFSLSAGRGQHPKEKITDR 66
OY 72 RRKY----GFIYKTHLFGPTVRVMGADNVRRILLGEHRLVSVHWPASVRTILGAGCLSN 127
Db 67 VRRFSSSSSCVFKTHLFGSPTAVVTGASGNKFLFTNENKLVSWWPDVSNKIFP---SS 122
OY 128 LHDSHKQKKKVIIM---QAFSREALOCYLVIAEVSACLE-QWLSGGERGLLVYPEVKR 183
Db 123 MQTSSKEARKLMLLSQFMKPEALRRYGVWDEIAQRHFETEW--ANQDVIVPPLTKK 180
OY 184 LMERIAMRILLGCEPBPAGGGEDEQQLVEAFEEEMTRNLFSLPIDVPFSGLYRGVKARNLI 243
Db 181 FTFSIACRSFLSME-DPA---RVRQLEEDFNTVAVGIFSIDLPGRFRNRAIKASRL 235
OY 244 HARIENIRAKIRIQAATEPDGCKDALOLLIEHSWGERLDMQALKOSTELLFGGHE 303
Db 236 RKEVSAIVRQKKEELKA---GKALEHDLISHMIMNIGETKD-EDLADKIIGLLIGHD 290
OY 304 TTASAATSLITYLGYPHVLQKVREIKSKGLCKSNQDNKLDMETLEOLKYIGCVIKET 363
Db 291 TASIVCTFVNVYLAEPHYQRLQE--QKEILKEKKEKEGLRWEDIEKMYSWNVACEV 348
OY 364 LRINPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHDVADIFTNKEEFNDRFITVPH 423
Db 349 MRIVPPLSGTFREALDHFSFKGYIIPKGMKLYWSATATHMNPDYFPEPERFENRFEESG 408
OY 424 PEDASRFSPFPGGLRSCVGEKEFAKILKIFTVELARHCDWQ-LLNGPPTMKTSPVTYP 482
Db 409 PKP---YTYVPFGGCGPRMGCKEYARLEILIFMHNLVNRFKWEKVPFNENKIYVDPLPIP 465
OY 483 VDNIPAR 489
Db 466 DKGLPIR 472

RESULT 8
O9PUG2 PRELIMINARY; PRT; 107 AA.
AC O9PUG2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYP26 (FRAGMENT).
GN CYP26.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Martinez-Ceballos E., Burdsal C.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF185266; AAD56546.1; -.
DR HSSP; P14779; 1B07.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 107 AA; 12009 MW; CA68BFB8D598BC474 CRC64;

Query Match 20.6%; Score 538; DB 13; Length 107;
Best Local Similarity 90.7%; Pred. No. 8.6e-37;
Matches 97; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 373 GFRVALKTFELNGYQIPKGMNVIYSICDTHDVADIFTNKEEFNDRFITVPHPEDASRFSE 432
Db 1 GFRVALKTFELNGYQIPKGMNVIYSICDTHDVADLFTDKDEFNPDRTFMSPSPEDSSRFSF 60
OY 433 IPFGGLRSCVGEKEFAKILKIFTVELARHCDWQLLNGPPTMKTSP 479
Db 61 IPFGGLRSCVGEKEFAKILKIFTVELARSCDWQLLNGPPTMKTSP 107

RESULT 9
O9FH76 PRELIMINARY; PRT; 463 AA.
AC O9FH76;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYTOCHROME P450.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20181125; Pubmed-10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:31-63(2000).
DR EMBL; AB020744; BAB10255.1; -.
SQ SEQUENCE 463 AA; 52366 MW; CCD17293F53F812 CRC64;

Query Match 19.8%; Score 517.5; DB 10; Length 463;
Best Local Similarity 29.3%; Pred. No. 3.4e-34;
Matches 143; Conservative 88; Mismatches 212; Indels 45; Gaps 11;

Query Match	19.28;	Score 501.5;	DB 10;	Length 482;
Best Local Similarity	28.28;	pred. No. 7.4e-33;		
Matches 140;	Conservative 94;	Mismatches 219;	Indels 43;	Gaps 13;

```

QY      4 PALASALCTFVLPPLLEFLAALKLMDLYCVSSRRSCALPLPGTMCFFFEGETLQMVLO 63
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      20 PALITLTIVVVVLLF-----KWLHWKEQR-----LRLPGSGMGLPYIGETLRLYTE 68
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
QY      64 R-RKFLQMKRRKRYGYFIKTHLFGRPYTRVVMGADNVRRILLGHEHRLVSVHMPASVRTILGA 122
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      69 NPNSFATRQNKYGDIFKTHILGCPCVMISSPEAARMVLSKAHLFKPTYPSPKERMIGP 128
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
QY      123 GCLSNLHDSHKQKKVIMAFSREALQCVLVIAEYVSSCLEQWLSCGERGLLVYPEVK 182
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      129 EALFFHQGPYHSTLKRVLQSSFMPSALRPYVSHIELLVLOTLSSWTSQKSINTLEY--MK 186
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
QY      183 RLMFRIAMRILLGCEPPAGGEDQQLVEA----FEEMTRNLFSLPIDVPEFSGLYRGVK 238
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      187 RYAFDVAIM-----SAFGDKEEPTTIDVIKLYQRLERGYNSMPLDLPGLTFHKSMM 238
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
QY      239 ARNLHARIEENIRAKI--RLQATEPDDGGCKDALQLLIEHSWGERLDMQALQOSTEL 297
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      239 AR---IELSEELRKVIEKRRNGREGG---LLGVLLGAKDQKRNGLSDSQIADNIIGV 291
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
QY      298 LFGGHETTASATSLITTYLGLPHYLOKV-REIKSKGLCKSNQDNKLDMEITLQLYKI 356
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      292 IFAATDTPTASVLTWLKYLHDHNPDLQEVSRQFSIRQIKKENR--RISWEDTRKMPLT 349
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
QY      357 GCVIKETLRNPVPVPGGFRVALKTFELNGYQIPKGMVNIYSICDTHDVADIFTNKEEFP 416
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      350 TRVIOETLRASVLSFTFREAVODVEYDGLIPKGWKLPLFRRIHSSSEFFPDPEKFD 409
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
QY      417 DRF-IVPHPEDASRFSFIPEGGGIRSCVGKEFAKILKIFTVELARHCDWOLLNGDPPTMK 475
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      410 SRFEVAPKP-----YTMPFGNGVHSCPGSEIAKLEMLILHLHTTSFRWEVIGDEGIG 464
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
QY      476 TSPTVYPVDNLPARFT 491
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      465 YGPFVPVKKGLPIRVT 480
      111: : : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :

RESULT 12
065624 PRELIMINARY; PRT; 457 AA.
ID 065624;
AC 065624;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME P450.
GN T18B16.200 OR AT4G19230.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-457 FROM N.A.
RA Van Der Schueren J., Chuang Y.J., Aert R., Defoor E., Robben J.,
RA Volckaert G., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.

```

RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; AL021687; CAA1613.1; -.
DR EMBL; AL161550; CAB78925.1; -.
DR HSSP; P33006; ICPT.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 457 AA; 52436 MW; 6105FB7C181E4F07 CRC64;

Query Match	18.9%;	Score 494.5;	DB 10;	Length 457;
Best Local Similarity	29.5%;	Pred. No. 2.6e-32;		
Matches 137; Conservative	92;	Mismatches 191;	Indels 45;	Gaps 13;

[illegible]

```

RESULT 13
064989
ID 064989 PRELIMINARY; PRT; 513 AA.
AC 064989;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE STEROID 22-ALPHA-HYDROXYLASE.
GN DMF4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WS-2;
RA Choe S., Dilkes B.P., Fujioka S., Takatsuto S., Sakurai A.,
RA Feldmann K.A.;
RL Plant Cell 10:0-0(1998).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF044216; AAC05093.1; -.

```

DR InterPro; IPR001128; -.
DR Pfam; PF00067; p450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58867 MW; B1639BDD9A4DA6F3 CRC64;

Query Match 17.6%; Score 461; DB 10; Length 513;
Best Local Similarity 28.4%; Pred. No. 1.7e-29;
Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

QY 3 LPALLASALCTFVLPPLFLAALKLMDLYCVSSRDRSCALPLPGTMGFPEFGETL----- 58
DB 10 LPLLLPSL-----LSLLFLILK-----RRNRKTRFNLPPGKSGWFLGETTIGYLK 57
QY 59 -QWVLQRRKFLQMKRRKKGFIYKTHLFGPRTVRVMGADNVRRIILGHRILVSVHWPASVR 117
DB 58 PYTATTLGDFWQOHVSKYGIYRSNLFGEPTIVSADAGLNRFILQNEGRLEFECSTYPRSIG 117
QY 118 TILGAGCLSNLHDSHKQKKVIMQAFSREALQCYLV-IAEEVSSCLEQWLSCGERGL 176
DB 118 GILGKWSMLVLVGDMDRMSISLNFISHARLRTILKDKVERHTLFVLDWQO--NSIFS 175
QY 177 VYPEVKRLMERIAMRILLGCEPAGGEDQQLVEAFEEMTRNLSLPIDVPFSGLYRG 236
DB 176 AQDEAKKFTFNIMAKHIMSMDPGE-----ETEQLKKEYVTFMKGVASAPLNLPGTAYHKA 231
QY 237 VKARNLIHARIEENIRAKIRLQATEPDGCKDALQLIEHSWE-----RGERLDMQ- 288
DB 232 LQSRATILKFIERKME--RKLDIKEED--QEEEVKTEDEAEMSKSDHVRKQRTDDDL 286
QY 289 ---ALKOS--STE-----LIFGHEHTTASATSLITYLGLYPHYLQKVREIKSKGL 335
DB 287 LGWVLKHSNLSTEQILDLSLTFAGHETSSVAIALAIFLQACPKAVEELREHLEIAR 346
QY 336 LCKSNODNKLDMETLEQLKYIGCVIKETRLNRPVPGGFVALKTFELNGYQIPKGMNVI 395
DB 347 AKKELGESELNWDYKKMDFQCVINETRLGNVRFHLRKALKDVRKYGYDIPSGWKVL 406
QY 396 YSICDTHDVADIFTNKEEFNDRFIVRPHEDASRFS-----FIFGGGLRSCVCKE 446
DB 407 PVISAVHLDNSRYDQPNLFNFWRQOQNNGASSSGSFSTWGNMNYMPFGGPRLCAGSE 466
QY 447 FAKILLKIFTVELARHCDWOLNRPPTMTKTSPTVYPVDNLP 487
DB 467 LAKLEMAVFIHHLVLKFNWELAE-----DDQPFAPFPVDFP 502

RESULT 14

Q9SCQ9 PRELIMINARY; PRT; 513 AA.
AC Q9SCQ9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE STEROID 22-ALPHA-HYDROXYLASE (DMF4).
GN T3A5.40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bioecker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RT EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -i- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AL132979; CAB62435.1; -.

DR InterPro; IPR001128; -.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; p450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58867 MW; B1639BDD9A5D7C93 CRC64;

Query Match 17.6%; Score 461; DB 10; Length 513;
Best Local Similarity 28.4%; Pred. No. 1.7e-29;
Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

QY 3 LPALLASALCTFVLPPLFLAALKLMDLYCVSSRDRSCALPLPGTMGFPEFGETL----- 58
DB 10 LPLLLPSL-----LSLLFLILK-----RRNRKTRFNLPPGKSGWFLGETTIGYLK 57
QY 59 -QWVLQRRKFLQMKRRKKGFIYKTHLFGPRTVRVMGADNVRRIILGHRILVSVHWPASVR 117
DB 58 PYTATTLGDFWQOHVSKYGIYRSNLFGEPTIVSADAGLNRFILQNEGRLEFECSTYPRSIG 117
QY 118 TILGAGCLSNLHDSHKQKKVIMQAFSREALQCYLV-IAEEVSSCLEQWLSCGERGL 176
DB 118 GILGKWSMLVLVGDMDRMSISLNFISHARLRTILKDKVERHTLFVLDWQO--NSIFS 175
QY 177 VYPEVKRLMERIAMRILLGCEPAGGEDQQLVEAFEEMTRNLSLPIDVPFSGLYRG 236
DB 176 AQDEAKKFTFNIMAKHIMSMDPGE-----ETEQLKKEYVTFMKGVASAPLNLPGTAYHKA 231
QY 237 VKARNLIHARIEENIRAKIRLQATEPDGCKDALQLIEHSWE-----RGERLDMQ- 288
DB 232 LQSRATILKFIERKME--RKLDIKEED--QEEEVKTEDEAEMSKSDHVRKQRTDDDL 286
QY 289 --ALKOS--STE-----LIFGHEHTTASATSLITYLGLYPHYLQKVREIKSKGL 335
DB 287 LGWVLKHSNLSTEQILDLSLTFAGHETSSVAIALAIFLQACPKAVEELREHLEIAR 346
QY 336 LCKSNODNKLDMETLEQLKYIGCVIKETRLNRPVPGGFVALKTFELNGYQIPKGMNVI 395
DB 347 AKKELGESELNWDYKKMDFQCVINETRLGNVRFHLRKALKDVRKYGYDIPSGWKVL 406
QY 396 YSICDTHDVADIFTNKEEFNDRFIVRPHEDASRFS-----FIFGGGLRSCVCKE 446
DB 407 PVISAVHLDNSRYDQPNLFNFWRQOQNNGASSSGSFSTWGNMNYMPFGGPRLCAGSE 466
QY 447 FAKILLKIFTVELARHCDWOLNRPPTMTKTSPTVYPVDNLP 487
DB 467 LAKLEMAVFIHHLVLKFNWELAE-----DDKPFAPFPVDFP 502

RESULT 15

Q9FMA5 PRELIMINARY; PRT; 465 AA.
AC Q9FMA5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME P450.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98290546; PubMed=9628582;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RT Sequence features of the regions of 1,456,315 bp covered by nineteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:41-54(1998).
DR EMBL; AB009048; BAB08653.1; -.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:39:40 ; Search time 62.8 Seconds
(without alignments)
271.098 Million cell updates/sec

Title: US-09-668-482-32
Perfect score: 2612
Sequence: 1 MGLPALLASALCTFVLPLLL.....PTVYPVDNLPRFTYFGDI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2612	100.0	497	1 CP26_MOUSE	O55127 mus musculu
2	2441	93.5	497	1 CP26_HUMAN	O43174 homo sapien
3	1734	66.4	492	1 CP26_BRARE	P79739 brachydanio
4	450	17.2	472	1 CP901_ARATH	Q42569 arabidopsis
5	442	16.9	464	1 CP85_LYCES	Q43147 lycopersico
6	442	16.9	490	1 CP883_ARATH	O23051 arabidopsis
7	410	15.7	492	1 C136_MYCTU	P95099 mycobacteri
8	359	13.7	519	1 CP881_MAIZE	Q43246 zea mays (m
9	339	13.0	503	1 CP51_RAT	Q64654 rattus norv
10	337	12.9	520	1 CP84_ARATH	Q42600 arabidopsis
11	332	12.7	503	1 CP51_PIG	O46420 sus scrofa
12	327.5	12.5	505	1 CP48_DROME	O9vs79 drosophila
13	326.5	12.5	493	1 C132_DROME	O9vgb4 drosophila
14	326	12.5	501	1 CP42_DROME	Q27589 drosophila
15	326	12.5	503	1 CP51_HUMAN	O16850 homo sapien
16	325	12.4	511	1 CP4C_BIADI	P29981 blaberus di
17	323	12.4	472	1 C13B_MYCTU	O53765 mycobacteri
18	322.5	12.3	476	1 C137_MYCTU	O69653 mycobacteri
19	322.5	12.3	511	1 CP45_RABIT	P14579 oryctolagus
20	317	12.1	441	1 C138_MYCTU	P96813 mycobacteri
21	314.5	12.0	512	1 CP4D1_DROME	P33269 drosophila
22	312	11.9	518	1 CP3R_ONCMY	O42563 oncorhynch
23	309	11.8	494	1 C134_DROME	O9vg40 drosophila
24	308.5	11.8	459	1 CPXN_ANASP	P29980 anabaena sp
25	307.5	11.8	511	1 CP47_RABIT	P14581 oryctolagus
26	307	11.8	483	1 C311_DROME	O9vyg7 drosophila
27	307	11.8	496	1 C7AE_DROME	O46054 drosophila
28	307	11.8	511	1 C772_SOLME	P37124 solanum mel
29	305.5	11.7	512	1 C4D1_DROSI	O16805 drosophila
30	305.5	11.7	519	1 CP4Y_HUMAN	O02928 homo sapien
31	305	11.7	500	1 CP46_MOUSE	O9vwb8 mus musculu
32	304.5	11.7	510	1 CP46_RABIT	P14580 oryctolagus
33	301.5	11.5	487	1 C135_DROME	O9vgb5 drosophila

34	301	11.5	520	1 CP33_HUMAN	Q08477 homo sapien
35	300.5	11.5	537	1 CP66_RAT	P51871 rattus norv
36	300	11.5	576	1 CP92_SOYBN	O48921 glycine max
37	299.5	11.5	509	1 CP41_RAT	P08516 rattus norv
38	298	11.4	500	1 CP46_HUMAN	O9y6a2 homo sapien
39	296.5	11.4	507	1 CP3S_BOVIN	P79102 bos taurus
40	293.5	11.2	512	1 CP51_CUNEL	O9uvc3 cunninghame
41	293	11.2	506	1 CP44_RABIT	P10611 oryctolagus
42	293	11.2	507	1 CP4E_DROME	O46051 drosophila
43	291.5	11.2	508	1 CP48_RAT	P24464 rattus norv
44	289.5	11.1	492	1 C133_DROME	O9vgb3 drosophila
45	289.5	11.1	506	1 CP4B_RABIT	P15128 oryctolagus

ALIGNMENTS

RESULT 1
ID CP26_MOUSE STANDARD; PRT; 497 AA.

AC 055127;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26 OR P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C3H;
RC MEDLINE=97392446; PubMed=9250660;
RX Fujii H., Sato T., Kaneko S., Gotoh O., Fujii-Kuriyama Y., Osawa K.,
RA Kato S., Hamada H.;
RA "Metabolic inactivation of retinoic acid by a novel P450
RT differentially expressed in developing mouse embryos.";
RL EMBO J. 16:4163-4173(1997).

[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=98113212; PubMed=9442090;
RX Abu-Abed S.S., Beckett B.R., Chiba H., Chithalen J.V., Jones G.,
RA Metzger D., Chambon P., Petkovich M.;
RT "Mouse P450RA1 (CYP26) expression and retinoic acid-inducible retinoic
RT acid metabolism in F9 cells are regulated by retinoic acid receptor
RT gamma and retinoid X receptor alpha.";
RL J. Biol. Chem. 273:2409-2415(1998).

-!- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -!- INDUCTION: BY RETINOIC ACIDS (RA).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Y12657; CAA73206.1; -
DR MGI: 1096359; CYP26.
DR InterPro: IPR001128; -
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;

KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56177 MW; 33B07D7C29134471 CRC64;

Query Match 100.0%; Score 2612; DB 1; Length 497;
Best Local Similarity 100.0%; Pred. No. 5.7e-175;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPPLLFALALKLMDLYCVSSRDRSCALPLPPTMGFPFPGETLQM 60
Db 1 MGLPALLASALCTFVLPPLLFALALKLMDLYCVSSRDRSCALPLPPTMGFPFPGETLQM 60
QY 61 VLQRRKFLQMKRRKYGFITYKTHLFGPRTVRVMGADNVRRILLGEHRLVSVHWPASVRTIL 120
Db 61 VLQRRKFLQMKRRKYGFITYKTHLFGPRTVRVMGADNVRRILLGEHRLVSVHWPASVRTIL 120
QY 121 GAGCLSNLHDSHKKQKKVIMQAFSREALQCYVLVIAEEVSSCLEQWLSCGERGLLVYPE 180
Db 121 GAGCLSNLHDSHKKQKKVIMQAFSREALQCYVLVIAEEVSSCLEQWLSCGERGLLVYPE 180
QY 181 VKRLMFIAMRILGCEPAGGGEDEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
Db 181 VKRLMFIAMRILGCEPAGGGEDEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
QY 241 NLIHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
Db 241 NLIHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
QY 301 GHETTASAATSLITYLGLYPHYLQKVREIKSKGLCKSNQDNKLDMETLEQLKYIGCVI 360
Db 301 GHETTASAATSLITYLGLYPHYLQKVREIKSKGLCKSNQDNKLDMETLEQLKYIGCVI 360
QY 361 KETLRLNPPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADIFTNKEEFNDRFI 420
Db 361 KETLRLNPPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADIFTNKEEFNDRFI 420
QY 421 VPHPEDASRSFIFPGGGLRSCVCKEFAKILKIFTVELARHCDWOLLNGPPTMKTSPV 480
Db 421 VPHPEDASRSFIFPGGGLRSCVCKEFAKILKIFTVELARHCDWOLLNGPPTMKTSPV 480
QY 481 YPVDNLPARFTYFQGD 497
Db 481 YPVDNLPARFTYFQGD 497

RESULT 2
CP26_HUMAN
ID CP26_HUMAN STANDARD; PRT; 497 AA.
AC 043174;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373542; PubMed=9228017;
RA White J.A., Beckett-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,
RA Jones G., Petkovich M.;
RT "cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RA1)
RT identifies a novel family of cytochromes P450.";
RL J. Biol. Chem. 272:18538-18541(1997);
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98380037; PubMed=9716180;
RA Sonneveld E., van den Brink C.E., van der Leede B.M., Schulkes R.K.,
RA Petkovich M., van der Burg B., van der Saag P.T.;

RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for
RT all-trans-RA and can be induced through RA receptors in human breast
RT and colon carcinoma cells.";
RL Cell Growth Differ. 9:629-637(1998).

RP TISSUE SPECIFICITY.
RX MEDLINE=99045433; PubMed=9826557;
RA Trofimova-Griffin M.E., Juchau M.R.;
RT "Expression of cytochrome P450RA1 (CYP26) in human fetal hepatic and
RT cephalic tissues.";
RL Biochem. Biophys. Res. Commun. 252:487-491(1998).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN ADULT LIVER, HEART,
CC PITUITARY GLAND, ADRENAL GLAND, PLACENTA AND REGIONS OF THE BRAIN.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF005418; AAB88881.1; -.
DR MIM: 602239; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56162 MW; EAB6B84B24B2EAB3 CRC64;

Query Match 93.5%; Score 2441; DB 1; Length 497;
Best Local Similarity 93.4%; Pred. No. 4.7e-163;
Matches 464; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPPLLFALALKLMDLYCVSSRDRSCALPLPPTMGFPFPGETLQM 60
Db 1 MGLPALLASALCTFVLPPLLFALALKLMDLYCVSSRDRSCALPLPPTMGFPFPGETLQM 60
QY 61 VLQRRKFLQMKRRKYGFITYKTHLFGPRTVRVMGADNVRRILLGEHRLVSVHWPASVRTIL 120
Db 61 VLQRRKFLQMKRRKYGFITYKTHLFGPRTVRVMGADNVRRILLGDRLVSVHWPASVRTIL 120
QY 121 GAGCLSNLHDSHKKQKKVIMQAFSREALQCYVLVIAEEVSSCLEQWLSCGERGLLVYPE 180
Db 121 GAGCLSNLHDSHKKQKKVIMQAFSREALQCYVLVIAEEVSSCLEQWLSCGERGLLVYPE 180
QY 181 VKRLMFIAMRILGCEPAGGGEDEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
Db 181 VKRLMFIAMRILGCEPQLAGDGDSEQQLVEAFEEEMTRNLSLPIDVPFSGLYRGVKAR 240
QY 241 NLIHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
Db 241 NLIHARIEQNIRAKICGLRASEAGGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
QY 301 GHETTASAATSLITYLGLYPHYLQKVREIKSKGLCKSNQDNKLDMETLEQLKYIGCVI 360
Db 301 GHETTASAATSLITYLGLYPHYLQKVREIKSKGLCKSNQDNKLDMETLEQLKYIGCVI 360
QY 361 KETLRLNPPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADIFTNKEEFNDRFI 420
Db 361 KETLRLNPPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVADIFTNKEEFNDRFI 420


```

QY 421 VPHPEDASRFSFIPFGGGLRSCVCGKEFAKILKIFIYELARHCDWOLLNGPPTMTKTSPTV 480
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 421 LPHPEDASRFSFIPFGGGLRSCVCGKEFAKILKIFIYELARHCDWOLLNGPPTMTKTSPTV 480

QY 481 YPVDNLPARFTYFOGDI 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 481 YPVDNLPARFTTHFGEI 497

RESULT 3
CP26_BRARE STANDARD; PRT; 492 AA.
ID CP26_BRARE P79739;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=97094702; PubMed=8939936;
RA White J.A., Guo Y.-D., Baetz K., Beckett-Jones B., Bonasoro J.,
RA Hsu K.E., Dilworth F.J., Jones G., Petkovich M.;
RT "Identification of the retinoic acid-inducible all-trans-retinoic
RT acid 4-hydroxylase."
RL J. Biol. Chem. 271:29922-29927(1996).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF 4-HYDROXYLATION; HYDROXYLATION.
CC RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED FORMS OF RA,
CC INCLUDING 4-OH-RA AND 4-OXO-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; U68234; AAC60045.1; -.
DR ZFIN; ZDB-GENE-990415-44; cyp26.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR OXidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 438 438 HEME (BY SIMILARITY).
SQ SEQUENCE 492 AA; 56281 MW; FD471435B2F30509 CRC64;

Query Match 66.4%; Score 1734; DB 1; Length 492;
Best Local Similarity 67.7%; Pred. No. 8.6e-114;
Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

QY 1 MGLPALLASALCTFVLPPLLEFLAALKLMDLYCVSSRRDRSCALPLPPTMGPFPGFTIOM 60
||| 1: : ||| |||:|||||:||||: : 1: ||| ||| ||| ||| ||| |||
Db 1 MGLITLMTVTLCTIVLPVLLEFLAALKWEMLMIRVDPNCRSPPLPPTMGLPFIGETIOL 60

QY 61 VLQRKFLQMKRRKRYGFITYKTHLFGRPYVRVMGADNVRRIILGHRLYSVHWPASVRTIL 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 ILQRKFLRMKRQYGCICYKTHLFGNPTVRVMGADNVRQIILGEHKLVSQWMPASVRTIL 120

```

[illegible]

```

RESULT      4
ID           C901_ARATH      STANDARD;          PRT;      472 AA.
AC           042569;
DT           15-DEC-1998 (Rel. 37, Created)
DT           15-DEC-1998 (Rel. 37, Last sequence update)
DT           15-DEC-1998 (Rel. 37, Last annotation update)
DE           CYTOCHROME P450 90A1 (EC 1.14.-.-).
GN           CYP90A1 OR CYP90 OR CPD.
OS           Arabidopsis thaliana (Mouse-ear cress).
OC           Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC           Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC           Brassicales; Brassicaceae; Arabidopsis.
OX           NCBI_TaxID=3702;
RN           [1]
RP           SEQUENCE FROM N.A.
RC           STRAIN=CV. COLUMBIA;
RX           MEDLINE=96200769; PubMed=8612270;
RA           Szekeres M., Nemeth K., Koncz-Kalman Z., Mathur J., Kauschmann A.,
RA           Altman T., Redei G.P., Nagy F., Schell J., Koncz G.;
RT           "Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
RT           controlling cell elongation and de-etiolation in Arabidopsis.";
RL           Cell 85:171-182(1996).
CC           -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC           -----
CC           This SWISS-PROT entry is copyright. It is produced through a collaboration
CC           between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC           the European Bioinformatics Institute. There are no restrictions on its
CC           use by non-profit institutions as long as its content is in no way
CC           modified and this statement is not removed. Usage by and for commercial
CC           entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC           or send an email to license@isb-sib.ch).
CC           -----
DR           EMBL; X87367; CAA60793.1; -.
DR           EMBL; X87368; CAA60794.1; -.
DR           InterPro; IPR001128; -.
DR           Pfam; PF00067; P450; 2.
DR           PROSITE; PS00086; CYTOCHROME_P450; 1.
KW           Oxidoreductase; Monooxygenase; Heme.
FT           BINDING      418      418      HEME (BY SIMILARITY).
SEQUENCE     472 AA; 53785 MW; 41A73F46D64E343F CRC64;

```


CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AC000098; AAB71462.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT TRANSMEM 6
FT BINDING 439 439 HEME (BY SIMILARITY).
SQ SEQUENCE 490 AA; 56409 MW; 7FD8CD7A8864D298 CRC64;

Query Match 16.98; Score 442; DB 1; Length 490;
Best Local Similarity 27.3%; Pred. No. 9.3e-24;
Matches 127; Conservative 84; Mismatches 217; Indels 38; Gaps 12;

QY 44 LPPGTGGEPPFEGTLLQWVLORR-----KFLQMKRRKRYG--FIYKTHLFGRPYVRMGADN 96
DB 45 LPPGDLGWPFIGNMLSFLRAFKTSDDPSFTRTLIKRYGPKGIYKAHMFNGNPSIITYTSDT 104
QY 97 VRRILGHRVSVHWPASVRLTGAGCLSNLHDSHKO--RKKVIMQAFSREALQCYLV 155
DB 105 CRRVLTDDDAF-KPGWPTSTMELIGRKSFGVGSFEEHKKRLRLTAAPVNGHEALSTIPIY 163
QY 156 IAEVSSCLEQWLSGGERGLVYPEVKRIMFRIAMRILGCEPAGGEGEQOLVAF 215
DB 164 IEEVNTVLDKWTMGEFEFLTH--LRKLTFRIMYIFLSE-----SENVMDALE 212
QY 216 E---MTRNLESLPIDVPFSGLYRGVAKARNLIHARIEENIRAKIRLQATEPPDGCKDAL 271
DB 213 REYALNKGVRAMAVNIPGFAYHRAFKARKTLVAAF-QSIYTERRNQKQNILSNKQDML 271
QY 272 QLLIEHSWGERLDMQALKQSTELLEFGHETTASATSLITYLGLYPHVLQKVREIK 331
DB 272 DNLNVKDEGCKTLDDDEIIDLVMYLNAGHSSGHTIMMATVFLQEHPEVLQRAKAE-- 329
QY 332 SKGLCKSNQDNK--LDMETLEQLKYGIVIKETRLRNPVPGFRVALKTFELNGYQIP 389
DB 330 -QEMILKSRPEGQKGLSLKETRKMEFLSQVDETFLRVITFSLTAFREAKTDVEMNGYLIP 388
QY 390 KGMNVITSCDTHDVADIFTNKEEFNDRF---IVPHPEDASRFSIFPGGGLRSCVGE 446
DB 389 KGWKVLTWERDVHIDPEVFPDPKRKDPARWDNGFVP-----KAGAFLLPGAGSHLCPGND 443
QY 447 FAKILKIFTVELARHCDWQLNGPPTMKTSPTVYPVDNLPAFTY 492
DB 444 LAKLEISIFLHHFLKLYQVKRSNPECPVMXLPHTPTDNCLEARISY 489

RESULT 7
C136_MYCTU STANDARD; PRT; 492 AA.
AC P95099;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PUTATIVE CYTOCHROME P450 136 (EC 1.14.-.-).
GN CYP136 OR RV3059 OR MTCY22D7.22C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gao S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z83866; CAB06263.1; -.
DR Tuberculist; RV3059; -.
DR InterPro; IPR001128; -.
DR InterPro; IPR002397; -.
DR InterPro; IPR002401; -.
DR InterPro; IPR002403; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PRINTS; PR00385; P450.
DR PRINTS; PR00463; EP450I.
DR PRINTS; PR00465; EP450IV.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
FT BINDING 439 439 HEME (BY SIMILARITY).
SQ SEQUENCE 492 AA; 56227 MW; B0A78FCE95622F3D CRC64;

Query Match 15.7%; Score 410; DB 1; Length 492;
Best Local Similarity 29.2%; Pred. No. 1.6e-21;
Matches 143; Conservative 78; Mismatches 186; Indels 82; Gaps 23;

QY 37 DRSCALPLPPT-----MGPFEGTLLQWVLORRKFLQMKRRKRYGFIYKTH---LFG 85
DB 42 EKRLAEP-PPSGSLKPVVDAGLPIIGHMIEMLRGSPDYLM-----FIYKTKGPVVF 93
QY 86 ----RPTVRVMGADNVRRIILGHRVSVH-WPASVRLTGAGCLSNLHDSHKKQKKVI 140
DB 94 DSAVLLPGVALGPDAAQVITYSNRKNKYSQQGWVPVIGPFEHNG-LMLLDFEEMFHRIM 152
QY 141 MQAFSREALQCYLVIAEYSSCL-EQWLSGGERGLVYPEVKRIMFRIAMRILGCEPG 199
DB 153 QEAFVRSRLAGYLEQMDRVVSRVAVDDWV-VNDARFLVYPAMKALTLDIASVFMGHEPG 211
QY 200 PAGGEGEQOLV---EAFEMTRN----LFSLPIDVPFSGLYRGVAKARNLIHARIEEN 250
DB 212 -----TDHELVTKVNKAFTITTRAGNAVIRTSVP---PFT-WWRGLRARELL-----EN 256
QY 251 ---IRAKIRRLQATEPPDGCKDALQLLIEHSWGERLDMQALKQSTELLEFGHETTAS 307
DB 257 YFTARVKERR---EASG--NDLLTVLCQTEDDGNRPSDADIVNHMIFLMAAHDISTS 310
QY 308 AATSLITYLGLYPHVLQKVREIKSKGLCKSNQDNKIDMETLEQLKYGIVIKETRLRN 367
DB 311 TATPMAYQLAHPWEQQRCDSEDRHG-----DGPLDIESLEQLSLDLYMNSIRLV 363
QY 368 PPVPGFRVALKTFELNGYQIPKGMNVITYSICDTHDVADIFTNKEEFNDRFIYVHPE-D 426
DB 364 TPVQWAMKQTVRDTELLGYLPRKGNVIAYPGMNHRLPELWTDPLTFDPERFTEPRNEHK 423
QY 427 ASRFSIFPGGGLRSCVGEKFAK-----LKFITVELARHCDWQLNGPPTMKTSP 478
DB 424 RHRYAFTPPGGGVHKICGMVFDQLEIKTILHRLRLRYRLELSRP-DYQ-----PRWDYSA 477
QY 479 TVYPVDNLP 487

```

Db      478  MPIPMGMP 486

RESULT      8
C881_MAIZE
ID      C881_MAIZE      STANDARD;      PRT;      519 AA.
AC      Q43246;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      CYTOCHROME P450 88A1 (EC 1.14.-.-) (DWARF3 PROTEIN).
GN      CYP88A1 OR D3.
OS      Zea mays (Maize).
OC      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC      Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC      Andropogoneae; Zea.
OX      NCBL_Taxid=4577;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV, B73;
RX      MEDLINE=96004534; Pubmed=7549486;
RA      Winkler R.G., Helentjaris T.;
RT      "The maize Dwarf3 gene encodes a cytochrome P450-mediated early step
RT      in gibberellin biosynthesis."
RL      Plant Cell 7:1307-1317(1995).
CC      -1- PATHWAY: EARLY STEP IN GIBBERELLIN BIOSYNTHESIS.
CC      -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, DEVELOPING LEAVES, THE
CC      VEGETATIVE MERISTEM, AND SUSPENSION CULTURE CELLS.
CC      -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL: U32579; AAC49067.1; -.
DR      InterPro: IPR001128; -.
DR      Pfam: PF00067; P450; 1.
DR      PROSITE: PS00086; CYTOCHROME_P450; 1.
KW      oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT      TRANSMEM      1      21      POTENTIAL.
FT      BINDING      466      466      HEME (BY SIMILARITY).
SQ      SEQUENCE      519 AA; 57906 MW; 0F8977A024316D95 CRC64;

Query Match      13.7%; Score 359; DB 1; Length 519;
Best Local Similarity 26.3%; Pred. No. 6.1e-18;
Matches 123; Conservative 89; Mismatches 213; Indels 42; Gaps 15;

QY      44  LPPGTMGFPFFEGETLQWVLQRRK-----FLQMKRRKRYG--FIYKTHLFGRPYVRVMGADN 96
      ||||| ||:| | : : : : : ||:| | : : : ||| | : |
Db      72  LPPGEMGWPLVGGMNAFLRAFKSGKPDAFIASFVRKRGTVGRYSFMSSPTVLVTTAEG 131
      ||||| :| | : : : : : ||:| | : : : ||| | : |
QY      97  VRRILGEHRLVSVHWEASVRTLIGAGCLSNLHDSHKORKKVMQAFSR-EALQCYLV 155
      ::||: : | : ||: : : ||: : : : : ||:| | : : : ||| | : |
Db      132 CKQVLMDDDAFVT-GWPKATVALVGPRSFVAMPYDEHRRIRKLTAPINGFDALTGYLPF 190
      ::||: : | : ||: : : ||: : : : : ||:| | : : : ||| | : |
QY      156 IAEVSSCLEQWLSGGERGLLVY-PEVKRLMFRIAMRILLGCEPPAGGEDE--QQLV 211
      | :| | | | | | : | : : : ||:| | : : : ||| | : |
Db      191 IDRTVTSSLRAWADHG--GSVEFELTELRRMTFKITVQIFL-----GGADQATRAL 240
      | :| | | | | | : | : : : ||:| | : : : ||| | : |
QY      212 EAFEEMTRNLFSLPIDVPFSGLYRG-VKARNLIHARIEENIRAKIRLQATEPDGCKDA 270
      ::| : : : : ||:| | : : ||:| | : : : ||| | : |
Db      241 RSYTELINYGMAAIIINLP-GFAYRGALRARRRLLVA-VLQGLVDERRAARAKVSGGVD 298
      | :| | | | | | : | : : : ||:| | : : : ||| | : |
QY      271 LQLLIHESWGERLDMQALKKOSTELLFGGHETTASAATSLITYLGLYPHYLQKREI 330
      :| | | | | | : | : : : ||:| | : : : ||| | : |
Db      299 MDRLEIAQDERGRHDDDEIIDLVMYLNAGHSSGHITMTMATVFLQENPDMFARAKAEQ 358
      | :| | | | | | : | : : : ||:| | : : : ||| | : |
QY      331 KSKGLCKSNQDNKLDMETLEQLKYGICVAKETLRLNPPVPGGFVALKTFELNGYQIPK 390

```

```

Db      359  EAIMRSIPSSQ-RGLTLRDRFRKMEYL$QVIDETLRLVNISFVSFRQATRDVFNQYLI PK 417
QY      391  GWNVIYSICDTHDVADIFTNKEEFNPDREFIVPHPEDASRFSFIPEGGLRSCVGEPAKI 450
Db      418  GWKQWLMYRSVHMDPQVYDPDTKFDPSRWEGHSFRAG---TFLAFGLGARLCPGNDLAKL 474
QY      451  LKIFT-----VELARHCDWQLLNGRPPTMKTSPYYPVDNLPARFT 491
Db      475  EISVFLHHLFLGYKLAR-----TNPRCRVRYLPHRPVNDCLAKIT 515

RESULT      9
CP51_RAT
ID      CP51_RAT      STANDARD;      PRT;      503 AA.
AC      064654; Q64549;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      CYTOCHROME P450 51 (EC 1.14.14.1) (CYPL1) (P450L1) (STEROL 14-ALPHA
DE      DEMETHYLASE) (LANOSTEROL 14-ALPHA DEMETHYLASE) (LDM) (P450-14DM).
GN      CYP51.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Liver;
RX      MEDLINE=94296405; PubMed=8024575;
RA      Aoyama Y., Funae Y., Noshiro M., Horiuchi T., Yoshida Y.;
RT      "Occurrence of a P450 showing high homology to yeast lanosterol 14-
RT      demethylase (P450(14DM)) in the rat liver.";
RL      Biochem. Biophys. Res. Commun. 201:1320-1326(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=WISTAR;
RX      MEDLINE=98158318; PubMed=9498553;
RA      Noshiro M., Aoyama Y., Kawamoto T., Gotoh O., Horiuchi T., Yoshida Y.;
RT      "Structural and evolutionary studies on sterol 14-demethylase P450
RT      (CYP51), the most conserved P450 monooxygenase: I. Structural analyses
RT      of the gene and multiple sizes of mRNA.";
RL      J. Biochem. 122:1114-1121(1997).
RN      [3]
RP      SEQUENCE OF 18-503 FROM N.A.
RC      TISSUE=Liver;
RX      MEDLINE=95394364; PubMed=7665087;
RA      Sloane D.L., So O.Y., Leung R., Scarafia L.E., Saldou N., Jarnagin K.,
RA      Swinney D.C.;
RT      "Cloning and functional expression of the cDNA encoding rat
RT      lanosterol 14-alpha demethylase.";
RL      Gene 161:243-248(1995).
CC      -I- FUNCTION: CATALYSES C14-DEMETHYLATION OF LANOSTEROL; IT TRANSFORMS
CC      LANOSTEROL INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL.
CC      -I- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC      OXIDIZED FLAVOPROTEIN + H(2)O.
CC      -I- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC      -I- SUBCELLULAR LOCATION: MICROSOMAL (POTENTIAL).
CC      -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D55681; BAA09529.1; -
DR      EMBL; AB004096; BAA20354.1; -
DR      EMBL; AB004087; BAA20354.1; JOINED.
DR      EMBL; AB004088; BAA20354.1; JOINED.
DR      EMBL; AB004089; BAA20354.1; JOINED.

```


CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U38416; AAC49389.1; -
DR EMBL; AL022141; CAA18128.1; -
DR EMBL; AF068574; AAD11580.1; -
DR EMBL; AL161589; CAB80293.1; -
DR InterPro; IPR001128; -
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Oxidoreductase; Monooxygenase; Membrane; Heme.
KW BINDING 458 458 HEME (BY SIMILARITY).
SQ SEQUENCE 520 AA; 58720 MW; E812779AF5BF01BC CRC64;

Query Match 12.9%; Score 337; DB 1; Length 520;
Best Local Similarity 25.8%; Pred. No. 2.1e-16;
Matches 133; Conservative 90; Mismatches 194; Indels 98; Gaps 23;

QY 15 VLPLLLFLAALKIMDLVCSSRRSCALPLPPGTMGPFEGETLQWLVLRKKFLQMKRRK 74
Db 22 VVSLFFIFIS-----FTRRRRP---PYPPGRGWPIIGNMLMMDQLTHRGLANLAKK 70

QY 75 YGFIYKTHL-FGRPTVRVMGADNVRRIILGHRLLVSVHWPASVRTILGACLSNL-HDSS 132
Db 71 YGGL--CHLWGLFLHMYAVSSPEVARQVLOVDVFSNRPAT-----AISLYTDRA 121

QY 133 -----HKQRKKVIMQAFSREALQCYLV-----IAEEVSSCLEQWLSCGERGL 175
Db 122 DMAFAHYGPPWRQMKVCVMKVSRRKRAESWASVRYDEVDMKVRVSCNVGKPINVGEO-- 179

QY 176 LVYPEVKRLMFRIAMRILLGCEPAGGGEDEQOLVEAFEEEMTR--NLESLPIDVPFSG 233
Db 180 -IFALTRNITYRAAF--SACEKG-----QDEFIRILQEEFSKLFCAFNVADFIPIYFGW 229

QY 234 Y-----RGVKARNLIHARIEENT--RAKIRLQATEPDDG-----CKDALQLIEHS 278
Db 230 IDPOGINKRLVKARNLDGFDIDIDEHMKKKENQNAVDDGDVDTMDVDDLAFYSEEA 289

QY 279 WERGERLDMQ-----ALKQSTELLFGHETTASAATSLITYLGLYPHVLQKVREI 330
Db 290 KLVSETADLQNSIKLTRDNIKAIIMDYMEGGTETVASAIEWALTELLRSPEDLKRVOQEL 349

QY 331 -KSKGLLCKSNQDNKLDMETLEQLKYGIVIKETRLNRPVPGGFRVALKTFELNGYQIP 389
Db 350 AEYVGL-----DRVVEESDIEKLTLYLCKTLKETLRMHPRIPILLHETAEDTSIDGFFIP 403

QY 390 KGMNVIYSICDTHVDADIFTNKEEFPNDRFIVPHPED--ASRFSFIPFGGGLRSCVGEF 447
Db 404 KSRVMINAFALGRDPTSWTDPDPTFRPSRFLPEGPVDFKGSNFEFIPFGSGRRSCPGMQ- 462

QY 448 AKILLKIFTVELA----RHC-DWQLLNGPPTMKT 477
Db 463 ----LGLYALDLAVAHILHCFWTKL---PDGMKPS 490

RESULT 11
CP51_PIG STANDARD; PRT; 503 AA.
AC 046420;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450 51 (EC 1.14.14.1) (CYPL1) (P450L1) (STEROL 14-ALPHA
DE DEMETHYLASE) (LANOSTEROL 14-ALPHA DEMETHYLASE) (LDM) (P450-14DM).
GN CYP51.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kojima M., Morozumi T., Hamasima N., Okamoto T.;
RT "Cloning of a pig lanosterol 14-demethylase cDNA."
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES C14-DEMETHYLATION OF LANOSTEROL; IT TRANSFORMS
CC LANOSTEROL INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: MICROSOMAL (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AB009988; BAA24134.1; -
DR InterPro; IPR001128; -
DR InterPro; IPR002403; -
DR Pfam; PF00067; p450; 2.
DR PRINTS; PR00465; EP450IV.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Transmembrane;
KW Heme; Cholesterol biosynthesis.
FT TRANSMEM 24 44 POTENTIAL.
FT BINDING 449 449 HEME (BY SIMILARITY).
SQ SEQUENCE 503 AA; 56866 MW; 0302949CE461AFD6 CRC64;

Query Match 12.7%; Score 332; DB 1; Length 503;
Best Local Similarity 23.2%; Pred. No. 4.4e-16;
Matches 117; Conservative 97; Mismatches 239; Indels 52; Gaps 14;

QY 6 LLASAL--CTFVPLLLFLAALKIMDLVCSSRRSCALPLPPGTMG-----PFFGE 56
Db 26 LLSSLLACAFLL-LLVYLFROAIGHL-----APLPAGAKSPPIYFSPPIPLGH 73

QY 57 TLQWLVQRKKFLQMKRRKYGFIYKTHLFGRPVVRVMGADNVRRIILGHRLLVSVH--WPA 114
Db 74 AIAFGKSPIEFLFNAYEKGYPVFSFTWVGKFTTYLLGSDAAALLFNKNEDLNAEDVYSR 133

QY 115 SVRTILGACLSNLHDSHKKQRKVIQAFSREALQCYLVIVIAEEVSSCLEQWLSCGERG 174
Db 134 LTTPVFGVAYDVNPVFLQKKMLKSGLNTAHFRQHVSIIIEKTKKEYFQSWGSEGERN 193

QY 175 LLVYPEVKRLMFRIAMRILLGCEPAGGGEDEQOLVEAFEEEMTRNL--FS-----LPI 226
Db 194 L--FEALSELLILTASHCL-----HGKEIRSQLINEKVAQLYADLDGGFSHAAMLPG 243

QY 227 DVPFSGLYRGVKARNLIHARIEENTIRAKIRRLQATEPDDGCKDALQLLIEHSWERGERLD 286
Db 244 WLPLPSFRR---RDRAHREIKNIFYKAIQKRROSEK--IDDLIQTLDDSTYKDGRLPT 297

QY 287 MQALKQSTELLFGHETTASAATSLITYLGLYPHVLQKVREIKSGLLCKSNQDNKLD 346
Db 298 DDEVAGMLIGLLLAGQHTSSTISAMWGFFLARDKTLQEKCYLEQKT---VCGEDLP-PLT 353

QY 347 METLEQLKYGIVIKETRLNRPVPGGFRVALKTFELNGYQIPKGMNVIYSICDTHVDAD 406
Db 354 YDQLKDNLNLDRCIKETLRLRPDIWMTMRMAKTPQTVAGYTIIPGHQVCVSPVYNQRLKD 413

QY 407 IFTNKEEFPNDRFIVPHPEDASRFSFIPFGGGLRSCVGEFAKILLKIFTVELARHCDWQ 466
Db 414 SWVERLDENPDRLQDNPASGEKFAVVPFAGGRHRCIGENFAVYQIKTIWSTMRLRYEFD 473

RX MEDLINE=20196011, PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 RA papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolelli J., Peter A., Schottler P., Werner M., Mourikioti F.,
 RA Belinert N., Dove G., Schafer U., Jackle H., Bucheton A.,
 RA Calister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
 RA McMillan P.J., Sales C., Tait E.A., Valenti P., Saunder R.D.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of D.
 RT melanogaster.";
 RL Science 287:2220-2222(2000).
 [31]
 RN SEQUENCE FROM N.A.
 RP STRAIN=BERKELEY;
 RC MEDLINE=20196006; PubMed=10731132;
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Burtis J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glödek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [4]
 RN SEQUENCE OF 4-501 FROM N.A.
 RP STRAIN=VARIOUS STRAINS;
 RC Phillips K.S., Begun D.J., Aquadro C.F.;
 RA "Evidence for non-neutral evolution around the cytochrome p450 gene
 RT cluster on the Drosophila melanogaster X chromosome.";
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN THE METABOLISM OF INSECT HORMONES AND IN
 CC THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
 CC OXIDIZED FLAVOPROTEIN + H(2)O.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC	-			
DR	EMBL; X75955;	CAA53568.1;	-.	
DR	EMBL; Z23005;	CAA80549.1;	-.	
DR	EMBL; AL009194;	CAAI5698.1;	-.	
DR	EMBL; AE003423;	AAF45741.1;	-.	
DR	EMBL; AF017006;	AAB71169.1;	-.	
DR	EMBL; AF017007;	AAB71170.1;	-.	
DR	EMBL; AF017008;	AAB71171.1;	-.	
DR	EMBL; AF017009;	AAB71172.1;	-.	
DR	EMBL; AF017010;	AAB71173.1;	-.	
DR	EMBL; AF017011;	AAB71174.1;	-.	
DR	EMBL; AF017012;	AAB71175.1;	-.	
DR	EMBL; AF017013;	AAB71176.1;	-.	
DR	EMBL; AF017014;	AAB71177.1;	-.	
DR	EMBL; AF017015;	AAB71178.1;	-.	
DR	EMBL; AF017016;	AAB71179.1;	-.	
DR	EMBL; AF017017;	AAB71180.1;	-.	
DR	EMBL; AF017018;	AAB71181.1;	-.	
DR	FlyBase; FBgn0011576;	Cyp4d2.		
DR	InterPro; IPR001128;	-.		
DR	InterPro; IPR002402;	-.		
DR	Pfam; PF00067;	P450; 1.		
DR	PRINTS; PR00385;	P450.		
DR	PRINTS; PR00464;	EP450II.		
DR	PROSITE; PS00086;	Cytochrome_P450; 1.		
KW	Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;			
KM	Endoplasmic reticulum; Polymorphism.			
FT	BINDING	449	449	HEME (BY SIMILARITY).
FT	VARIANT	163	163	K -> M (IN STRAINS CAM-44, CAM-48 AND BERKELEY).
FT	CONFLICT	2	25	MISSING (IN REF. 2).
FT	CONFLICT	30	30	I -> A (IN CAA80549).
FT	CONFLICT	160	160	A -> R (IN REF. 1).
FT	CONFLICT	487	501	LRSANGVHLGLKPRA -> GGRPTAFILA (IN REF. 1).
SQ	SEQUENCE	501 AA;	57914 MW;	8F8D98DBF39FDIA2 CRC64;

Query Match	12.5%;	Score 326;	DB 1;	Length 501;
Best Local Similarity	25.1%;	Pred. No. 1.2e-15;		
Matches 123; Conservative	92;	Mismatches 220;	Indels 56;	Gaps 15

OY	18	L L F L A A K L M D C Y V S R D S C A L P L P G T M B P E F F E T L O M - - - - V L O R R X F L O M K R	73
D b	9	L L V A F A T L L M D F L W - - - - R R G N G I L P G R P L D F L G N L L M Y R G L D P E Q I M D E V K K N O R	63
OY	74	K Y G F I Y K T H L F G R P T V Y R V M G A D N V R I L L G E H R L V S W H P A S V R T I L G A G L S N L D S H	133
D b	64	K Y G R L Y R W I L H O L A V F S T D P R D I E F V L S S Q O H I T K N N L Y K L L N C W L G D L M S T G R K W H	123
OY	134	K O R K K V I M O A F S R E A L O C Y V L V I A E V S S C L E O W L S - - G E R G L L Y P E V - - - - K R L M E R	187
D b	124	- G R R K I I T P T F H F K I L E Q V E I F D O O S A V M V E O L O S R A D G K T P I N I F P V I C L T A L D I I A E	182
OY	188	I A M R I L L G C E P P A G G E D E Q O L V E A F E E M T R N L F S - - - - L P I D V P F - - - - S G L Y R G	236
D b	183	T A M G T K I N A O K N P - - - - N L P Y V O A V N D V T N I L I K R F I H A M O R V D W I F R L T O P T E A K R Q	236
OY	237	V K A R N L I H A R I E E N I R A K - - - - I R L Q A T E P D G - - - - G C K D A L O L L - - I E H S W E R G E R L	285
D b	237	D K A I K V M H D F T E N I I R E R R E T L V N N S K E T T P E E V N F I G Q K R R M A L L D V L L Q S T I D G A P L	296
OY	286	D M O A L K O S T E L L F G H E T T A S A T S L I T Y I G L Y P H V L O K V R E I K S K G L L C K S N O D N K L	345
D b	297	S D E D I R E E V D T F M F E G H D T T S A I S F C L Y E I S R H P E V Q O R L Q O E I R D - - - V L G E D R K S P V	353
OY	346	D M E T L E Q L K Y I G C V I K E T L R N P P V P G G F R V A L K T F E L N G Y O I P K G W N V I Y S I C D T H V A	405

```

Db      354  TLRLDGLKLMENVIKESLRLHPVPMIGRWEAEVDVEIRGKHIPAGTNTMGIFVLLRDP 413
QY      406  DIETNKEEFNDRF--IVP--HPEDASRFSFIPEGGLRSCVCGKEFAKILKIFVELAR 461
      : | : | | | : | | | | : | | | | | : | | : | |
Db      414  EYFESPDEFERFERDADVPQIH-----YAYIPFSAGPRNCIGQKFAMLEMKSIVSKLLR 468
QY      462  HCDWQLNGPP 472
      | : | |
Db      469  HFELLPLGPDP 479

RESULT  15
CP51_HUMAN
ID      CP51_HUMAN          STANDARD;          PRT;          503 AA.
AC      Q16850; Q16784; O00770; O00772; Q99868;
DT      01-NOV-1997 (Rel. 35, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      CYTOCHROME P450 51 (EC 1.14.14.1) (CYPL1) (P450L1) (STEROL 14-ALPHA
DE      DEMETHYLASE) (LANOSTEROL 14-ALPHA DEMETHYLASE) (LDM) (P450-14DM).
OS      CP51.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Liver;
RC      MEDLINE=96201125; PubMed=8619637;
RX      Stromstedt M., Rozman D., Waterman M.R.;
RT      "The ubiquitously expressed human CYP51 encodes lanosterol 14 alpha-
RT      demethylase, a cytochrome P450 whose expression is regulated by
RT      oxysterols.";
RL      Arch. Biochem. Biophys. 329:73-81(1996).
[2]
SEQUENCE FROM N.A.
TISSUE=Liver;
RC      Noshiro M.;
RL      Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
TISSUE=Liver;
RC      MEDLINE=96404948; PubMed=8809088;
RX      Rozman D., Stromstedt M., Waterman M.R.;
RT      "The three human cytochrome P450 lanosterol 14 alpha-demethylase
RT      (CYP51) genes reside on chromosomes 3, 7, and 13: structure of the
RT      two retrotransposed pseudogenes, association with a line-1 element,
RT      and evolution of the human CYP51 family.";
RL      Arch. Biochem. Biophys. 333:466-474(1996).
[4]
SEQUENCE FROM N.A.
MEDLINE=97131516; PubMed=8975714;
RX      Rozman D., Stromstedt M., Tsui L.-C., Scherer S.W., Waterman M.R.;
RT      "Structure and mapping of the human lanosterol 14alpha-demethylase
RT      gene (CYP51) encoding the cytochrome P450 involved in cholesterol
RT      biosynthesis; comparison of exon/intron organization with other
RT      mammalian and fungal CYP genes.";
RL      Genomics 38:371-381(1996).
-1- FUNCTION: CATALYZES C14-DEMETHYLATION OF LANOSTEROL; IT TRANSFORMS
CC      LANOSTEROL INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL.
CC      CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC      OXIDIZED FLAVOPROTEIN + H(2)O.
-1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
-1- SUBCELLULAR LOCATION: MICROSOMAL (POTENTIAL).
CC      TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED WITH HIGHEST LEVELS IN
CC      TESTIS, OVARY, ADRENAL, PROSTATE, LIVER, KIDNEY, AND LUNG.
-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

```

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@lsb-sib.ch).
CC
CC -----
DR  EMBL; U23942; AAB39951.1; ALT_INIT.
DR  EMBL; D55653; BAA09512.1; ALT_INIT.
DR  EMBL; AC000120; AAB46356.1; ALT_INIT.
DR  EMBL; U51692; AAC50951.1; -.
DR  EMBL; U51684; AAC50951.1; JOINED.
DR  EMBL; U51685; AAC50951.1; JOINED.
DR  EMBL; U51686; AAC50951.1; JOINED.
DR  EMBL; U51687; AAC50951.1; JOINED.
DR  EMBL; U51688; AAC50951.1; JOINED.
DR  EMBL; U51689; AAC50951.1; JOINED.
DR  EMBL; U51690; AAC50951.1; JOINED.
DR  EMBL; U51691; AAC50951.1; JOINED.
DR  MIM; 601637; -.
DR  InterPro; IPR001128; -.
DR  InterPro; IPR002403; -.
DR  Pfam; PF00067; P450; 2.
DR  PRINTS; PR00465; EP450IV.
DR  PROSITE; PS00086; CYTOCHROME P450; 1.
KW  Oxidoreductase; Monooxygenase; Electron transport; Transmembrane;
KW  Heme; Cholesterol biosynthesis.
FT  TRANSMEM 24 44 POTENTIAL.
FT  BINDING 449 449 HEME (BY SIMILARITY).
FT  CONFLICT 272 272 R -> T (IN REF. 4).
FT  CONFLICT 368 368 K -> R (IN REF. 4).
SQ  SEQUENCE 503 AA; 56805 MW; 4FCCEC147FB4DED86 CRC64;

```

Query Match	12.5%	Score 326	DB 1	length 503
Best Local Similarity	23.5%	Pred. No. 1.2e-15		
Matches 119	Conservative 94	Mismatches 240	Indels 54	Gaps 13
QY 3	LPALLASALCTFVLPLLLFLAALKLMDLYCVSSRDRSCALPLPPTMG-----FPFFG	55		
Db 27	LSMLLIACAFLLSLVYLIRLAAGHL-----VOLPAGVKSPPYIFSPPIPLG	72		
QY 56	ETLQWVLQRRKFLQMKRRKYGFYKTHLFGKPTVRVMGADNVRILLGEHRLVSVH--WP	113		
Db 73	HALAFGKSPIEFLENAYEKYGPVFSFTWVGKTFYLLGSDAALLFNSKNEDLNAEDVYS	132		
QY 114	ASVFTILGAGCLSNLHDSHKQKKVIMQAFSREALQCYVLVIAEEVSSCLEQWLSCGER	173		
Db 133	RLTTPVEFGKVAYDVPNPVLEQKKMLKSLGINIAHFQHSIIEKETKEYFESWGEGEK	192		
QY 174	GLIYYPEVKRLMERIAMRIILGCEPGPAGGDEQQLVEAFEMTRNL--FS-----LP	225		
Db 193	N--VEEALSELLILTASHCL-----HGKEIRSQLENKYAQLYADLDGGFSHAWLLP	242		
QY 226	IDVPEFSGLYRGVKARNLIHARTEENI-RAKIRRLQATEPDPGGCKDALQLLIEHSWGER	284		
Db 243	GMPLPSPFRR---RDRAHREIKDIFYKAIQKRRQSEK---IDDILQTLDATYKDGPR	295		
QY 285	LDMQALKQSSTELLFGGHETTASATSLITYLGLYPRHVLQVREEIKSGKLCKSNQDNK	344		
Db 296	LTDDEVAGMLIGLLLAGQHTSSTTSAMWGFFLARDKTLQKKCYLEQKT--VCGENLP-P	351		
QY 345	LDMETLEQLKYIGCVIKETRLRNPVPVPGGFRVALKTFELNGYQIPKGNVITYSICDTHDY	404		
Db 352	LTYDQLKDLNLLDRCIKETRLRNPIMIMMRMARTPQVAGYTIIPGHQVCVSPYVQRL	411		
QY 405	ADIFTNKEEFNDRFTVPHPEDASRSFPIFGGGLRSVQKEFAKILLKIFTVELARHCD	464		
Db 412	KDSWVERLDENPDRLYLQDNPASGEKFAVYVPFAGGRHRCIGENFAYVQIKTIWSTMRLYE	471		
QY 465	WQLLNGP-PIMKTSPIVYPVDNLPARF	490		
Db 472	FDLIDGYEPTVNTYTMITHPENPVIRY	498		

OM of: US-09-668-482-32 to: GenEmbl:* out_format : pfs
Date: Nov 6, 2001 3:58 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09668482/runat_05112001_133353_10884/app-query.fasta_1.1680
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09668482_@cgn1_1_8677 -NCPD=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-668-482-32

Query length: 497

Database: GenEmbl:*

Database sequences: 1344157

Database length: -856060004

Search time (sec): 5019.990000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_rol:MMP450RA	+ 2612.00	3931.28	9.7e-211	1756	! Y12657 M.musculus mRNA for P45
gb_rol:AF115769	+ 2609.00	3926.92	1.7e-210	1726	! AF115769 Mus musculus cytochr
gb_pr4:AF005418	+ 2441.00	3672.91	2.4e-196	1743	! AF005418 Homo sapiens retinoid
gb_ov:AF199462	+ 2057.50	3094.96	3.7e-164	1479	! AF199462 Gallus gallus retinoid
gb_htg21:AL358613	+ 1939.50	2868.60	1.5e-151	166337	! AL358613 Homo sapiens chromo
gb_ov:AF057566	+ 1758.50	2643.06	5.5e-139	1479	! AF057566 Xenopus laevis retin
gb_ov:DRU68234	+ 1734.00	2606.04	6.4e-137	1479	! U68234 Danio rerio all-trans-1
gb_pr5:AF252297	+ 1048.50	1558.80	1.4e-78	4445	! AF252297 Homo sapiens cytochr
gb_ba3:SYCSLE	+ 669.50	953.08	7.5e-45	113064	! D64003 Synechocystis sp. PC
gb_ov:AF1855266	+ 538.00	813.94	4.2e-37	322	! AF185266 Gallus gallus CYP26 (C
gb_p12:AF318211	+ 496.50	735.61	9.7e-33	1494	! AF318211 Taxus cuspidata 5-alf
gb_p11:AC006931	+ 460.00	636.17	3.4e-27	116205	! AC006931 Arabidopsis thalian
gb_pat1:AR074363	+ 450.00	664.58	8.8e-29	1608	! AR074363 Sequence 1 from paten
gb_p12:ATCYP450R	+ 450.00	664.58	8.8e-29	1608	! X87367 A.thaliana mRNA for cyt
gb_p14:SLU54770	+ 442.00	653.94	3.4e-28	1395	! U54770 Solanum lycopersicum cy
gb_p12:AF318500	+ 442.00	653.29	3.7e-28	1487	! AF318500 Arabidopsis thaliana
gb_p12:AF279252	+ 425.00	626.54	1.2e-26	1649	! AF279252 Vigna radiata cytochr
gb_p11:AB008097	+ 418.00	614.34	5.5e-26	1934	! AB008097 Arabidopsis thaliana
gb_ba3:MYCY22D7	+ 410.00	573.76	1.0e-23	31859	! 283866 Mycobacterium tubercul
gb_p12:AF318501	+ 406.00	597.28	4.9e-25	1740	! AF318501 Arabidopsis thaliana
gb_p11:AB018112	+ 401.50	552.84	1.5e-22	70475	! AB018112 Arabidopsis thaliana
gb_p12:AF212991	+ 400.50	529.10	1.4e-24	1718	! AF212991 Cucurbita maxima ent-
gb_pr2:AC007002	+ 389.00	526.94	4.1e-21	140331	! AC007002 Homo sapiens BAC cl
gb_ba3:MXTRAH	+ 380.00	558.33	7.3e-23	1681	! AJ232955 Myxococcus xanthus cy
gb_p11:AB020744	- 374.00	512.57	2.6e-20	62052	! AB020744 Arabidopsis thaliana
gb_p12:AF326277	+ 368.50	540.01	7.6e-22	1844	! AF326277 Hordeum vulgare ent-k
gb_p13:ATT18B16	+ 365.50	495.23	2.4e-19	96574	! AL021687 Arabidopsis thaliana
gb_p12:ATCHRTV50	+ 365.50	487.86	6.1e-19	199382	! AL161550 Arabidopsis thaliana
gb_p14:ZMU32579	+ 359.00	526.34	4.4e-21	1724	! U32579 Zea mays DWARF3 (dwarf3
gb_p12:AF139532	+ 358.50	524.69	5.5e-21	1883	! AF139532 Liquidambar styracifl
gb_rol:AF166266	+ 351.00	513.64	2.3e-20	1830	! AF166266 Mus musculus lanoster
gb_p12:AF214009	+ 350.50	512.86	2.5e-20	1835	! AF214009 Brassica napus cytoch
gb_p12:AF214007	+ 342.00	499.76	1.3e-19	1880	! AF214007 Brassica napus cytoch
gb_p12:AF214008	+ 341.00	498.23	1.6e-19	1884	! AF214008 Brassica napus cytoch
gb_rol:RATCP14DM	+ 339.00	493.36	3.0e-19	2260	! D55681 Rattus sp. mRNA for cyt
gb_rol:RNUT17697	+ 339.00	492.82	3.3e-19	2383	! U17697 Rattus norvegicus lanos
gb_pat1:AR085702	+ 337.00	492.44	3.4e-19	1838	! AR085702 Sequence 1 from paten
gb_p13:ATU38416	+ 337.00	492.44	3.4e-19	1838	! U38416 Arabidopsis thaliana f
gb_p12:AF150881	+ 336.00	490.27	4.5e-19	1960	! AF150881 Lycopersicon esculent
gb_om:AB009988	+ 332.00	482.61	1.2e-18	2297	! AB009988 Sus scrofa mRNA for 1
gb_ba3:MTV025	- 329.00	437.76	3.8e-16	121125	! AL022121 Mycobacterium tuber

gb_pr10:HUMUL14D	+ 326.00	470.55	5.7e-18	3085	! D55653 Human mRNA for lanos
gb_pr10:HSU23942	+ 326.00	470.26	5.9e-18	3172	! U23942 Human lanosterol 14-
gb_rol:AB056457	+ 325.50	470.43	5.7e-18	2897	! AB056457 Mus musculus CYP4v
gb_in2:BLBP450A	+ 325.00	471.72	4.9e-18	2370	! M63798 Cockroach cytochrome

seq_name: gb_rol:MMP450RA

seq_documentation_block:

LOCUS MMP450RA 1756 bp mRNA ROD 08-JAN-1998
DEFINITION M.musculus mRNA for P450RA protein.
ACCESSION Y12657
VERSION Y12657.1 GI:2765213
KEYWORDS P450RA gene.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 1756)
Fujii,H., Sato,T., Kaneko,S., Gotoh,O., Fujii-Kuriyama,Y.,
Osawa,K., Kato,S. and Hamada,H.
Metabolic inactivation of retinoic acid by a novel P450
differentially expressed in developing mouse embryos

JOURNAL
2 (bases 1 to 1756)
Unpublished

REFERENCE
AUTHORS Hamada,H.
TITLE Direct Submission
JOURNAL Submitted (17-APR-1997) H. Hamada, Osaka University, Institute for
Mol. & Cell. Biol., 1-3 Yamada-oka, Suita, Osaka, 565, JAPAN

FEATURES
source
1. 1756
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/cell_line="P19"
/gene="P450RA"
57. 1550
/gene="P450RA"
/function="metabolises retinoic acid"
/codon_start=1
/product="P450RA protein"
/protein_id="CAA73206.1"
/db_xref="GI:2765214"
/db_xref="SWISS-PROT:O55127"
/translation="MGLPALLASALCTFVLLFLAALKLMDLYCVSSRDRSCALPL
PPTMGFPFGETLQWLQRRKFLQMKRRKGFYKTHLFGRPYRVMGADNVRILL
GEHRLVSVHWPASVRTILGAGCLSNLHDSHKQKKVIMQAFSREALOCYLVIAEEV
SSCLEQWLSCGERGLVYPEVKRIMFRIMRILICEPGAGGEDQQLVEAFEEEMT
RNLESRLIDVPEFSGLYRGVAKRLIMFRIAMRILICPEPGAGGEDQQLVEAFEEEMT
HSWGERGLDMQALKOSTELLFGHETTPASATSLITYGLYHYLVQKVRREIKSKG
LLCKSNQDNKLMDLETLQKLYIGCVIKETLRNPVPGGFRVALKTFELNGYQIPKGM
NVIVSICDTHVDADIFTNKEENPDREIVPHEDASRFSEIFPGGGLRSCVGEFAKI
LKIFTEVLARHCDWQLLNGPPTMTKTSPTVYVVDNLPAFTYFQGD1"

BASE COUNT 408 a 434 c 494 g 420 t
ORIGIN

alignment_scores:
Quality: 2612.00 Length: 497
Ratio: 5.256 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-668-482-32 x MMP450RA ..

Align seg 1/1 to: MMP450RA from: 1 to: 1756

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
57 ATGGGGCTCCCGGCTGCTGGCCAGTGCCTCTGCACCTTCGTGCC 106
17 OleuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrcysVal 34
|||||

107 GCTGCTGCTCTTCCTGGCGGCGCTCAAGCTCTGGACCTGTACTGTGTGA 156
34 erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
157 GCAGCCGCGATCGCAGCTGCCCTCCCTTGCCCCCGGTACCATTGGGC 206
51 phePropheheGlyIuThrLeuGlnMetValLeuGlnArgArgLysPh 67
207. TTCCCATTTCTTGGGAACAATTGCAGATGGTGTGCTTCAGCGAGAAGTT 256
67 eleuGlnMetLysArgArgLysTyrGlyPheIleYrLysThrHisLeuP 84
257 TCTGCAGATGAAGCGCAGGAATAAGGCTTCATCTACAGACGCATCTGT 306
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
307 TTGGGGGGCCCCACGGTGGCGGTGATGGCGCGGATATGTGCGGCGCATC 356
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
357 TTGCTGGGAGAGCACCGGTTGGTGTGCGTGCACCTGGCCCCGCGGTGCG 406
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
407 CACCATCTGGGGCGCTGGCTGCCCTCTCCAACTGCACGATTCCTCGCACA 456
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
457 AGCAGCGAAAGAGGTGATTTATGACAGCCCTCAGCCCGCAGGCACTCCAG 506
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGlnTr 167
507 TGCTACGTGCTCGTGTATCGCTGAGGAAGTCAGCAGTTGTCTGAGCAGTG 556
167 pleuSerCysGlyIuArgGlyLeuLeuValTyrProGluValLysArgL 184
557 GCTAAGCTGCGGCGAGCGCGGCTCTCTGCTACCCCGAGGTGAAGCGCC 606
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
607 TCATGTTCGCGATCGCCATGCGCATCTGCTGGGCTGCCAGCGGGTCCA 656
201 AlaGlyGlyGlyIuAspGluGlnGlnLeuValGluAlaPheGluGluMe 217
657 GCGGGCGGCGGGGAGGACGAGCAACAGCTCTGTGAGAGGCTTTCGAGGAGAT 706
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
707 GACCCGCAATCTCTTCTCTCTCCCATTGACGTGCCCTTTAGCGGCTGT 756
234 yTrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
757 ACCGGGGCGGTGAAGCGCGGAACTTATACAGCGCGCATCGAGGAGAAC 806
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
807 ATTGCGCGCCAGATCCGCGGCTTCAGGCTACAGACGCGGATGGGGTTG 856
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerThrGluLeuLeuPheGly 300
857 CAAGGACGCGCTGCAGCTCTGATGTGACACTCGTGGGAGAGGGGAGAGA 906
284 rGluAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
907 GGCCTGGATATGCAGGCACTAAACAATCTCAACAGAGCTCCTCTTTGGT 956
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
957 GGTCAATGAATACTACAGCCAGTGTCTGCACATCACTGATCACTTACTAGG 1006
317 yLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLysG 334
1007 ACTTACCACCATGTCTCTCAGAAAGTTGAGAAAGAGATAAAGAGCAAGG 1056

334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspmetGluThrLeu 350
1057 GCTTACTTTGCAAGAGCAATCAAGACAACAAGTAGACATGGAACCTTGG 1106
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
1107 GAACAGCTTAATACATATGGGTGTGCATTAAGAGAGACCTCGGATTGAA 1156
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
1157 TCCTCCGGTCCACAGAGGGTTCCGGGTGCTCTGAAGACTTTTGAAGCTGA 1206
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleYrSerIleCysAsp 400
1207 ATGGATACCAGATCCCAAGGGCTGGAATGTTATTATACAGTATCTGTGAC 1256
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
1257 ACCCAGCATGTGGCAGATATCTTCACTAACAAAGAGGAATTTAATCCCGA 1306
417 ParGpheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
1307 CCGCTTATAGTGCCCTCAATCCAGAGGATGCTTCCCGGTTCAGCTTCATT 1356
434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1357 CATTTGGAGAGGCGCTTCGAGCTGTGTAGGCAAAAGATTGCAAAAATT 1406
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
1407 CTCTTAAGATATTATACAGTGAGAGCTGGCTAGGCACGTGATTGGCAGCT 1456
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
1457 TCTAATGAGACCTCCTACAAATGAAGACAAGCCCACTGTGTACCCCTGTGG 1506
484 sPAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
1507 ACAATCTCCCTGCAAGATTCACTACTTCCAGGAGCATATATC 1547
seq_name: gb_rol:AF115769
seq_documentation_block: 1726 bp mRNA ROD 03-MAR-1999
LOCUS AF115769
DEFINITION Mus musculus cytochrome P450 retinoic acid metabolizing enzyme
ACCESSION AF115769
VERSION AF115769.1 GI:4324987
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1726)
Paine,C.T., Paine,M.L. and Sneed,M.L.
IDENTIFICATION of amelogenin- and tuftelin-interacting proteins using the yeast two-hybrid system
Connect. Tissue Res. 38, 257-267 (1998)
REFERENCE 2 (bases 1 to 1726)
Paine,C.T., Paine,M.L. and Sneed,M.L.
Direct Submission
Submitted (21-DEC-1998) CMDB, University of Southern California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
JOURNAL TITLE Location/Qualifiers
JOURNAL Direct Submission
FEATURES
source 1..1726
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="tooth"
/dev_stage="1 day old"
27..1520
/note="tuftelin-interacting protein 24 TIP-24"
/codon_start=1
CDS

/product="cytochrome P450 retinoic acid metabolizing
enzyme P450RA"
/protein_id="AAD17217.1"
/db_xref="GI:4324988"
/translation="MGIPALATATLCTFVLPLLFLAALKLMDLYCVSSRDRSCALPL
PGTMGFPEGETLQVLQRRKFLQMKRKYGYFYKTHLFGRPYRVMGADNVRILL
GEHRLVSVMWASVKTILGAGCLSNLHDSHKQKKVIMQAFSREALQCYLVLIAEEV
SSCLEQWLSCGERGLLVYPEVKRLMERIAMRILLGCEPAGGGEQQLVEAFEEMT
RNLFSLPIDVPSGLYRGVKARNLIHARIEENIRAKIRLQATEPDGCKDALQLIE
HSMGERLDMQALQKOSTELLFGHETTASATSLITYGLYPHYLOKVRREIKSKG
LLCKSNQDNKLDMEETLEOLKYIGCVIKEFTLRNPVPGFRVALKTFELNGYQIPKGM
NVYISICDTHVDADIFTNKEEFNPDRFIVPHPEDASRFSFIPEGGLRSCVKEFAKI
LKIFTVELARHCDWQLLNGPPTMKTSPTVYPVDNLPARFTYFQGD1"

BASE COUNT 405 a 428 c 476 g 417 t
ORIGIN

alignment_scores:
Quality: 2609.00 Length: 497
Ratio: 5.249 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.799

alignment_block:
US-09-668-482-32 x AF115769 ..

Align seg 1/1 to: AF115769 from: 1 to: 1726

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
27 ATGGGGCTCCCGCGCGCTGCTGGCCACTGCGCTCTGCACCTTCGCTGCC 76
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
|||||
77 GCTGCTGCTCTTCTGGCGCGCTCAAGCTCTGGGACCTGTACTGTGTGA 126
34 erSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
127 GCAGCGCGCATCGCACTGCGCCCTCCCTGCCCCCGGTACCATGGGC 176
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgLysPh 67
|||||
177 TTCCCATTTCTTGGGAAACATTGCAGATGTGCTTCAAGCGGAGGAAATT 226
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
227 TCTGCAGATGAAGCGCAGGAATAACGGCTTCATCTACAAGACGCATCTGT 276
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
277 TTGGGGCGGCCACCGGTGGGTGATGGCGCGGATATGTGGCGGCATC 326
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
327 TTGCTGGGAGAGCACCGGTGGTGTGCTGACCTGGCCCCCGCTCGGTGCG 376
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
377 CACCATCTCTGGCGCTGCTGCTCTCCAACTGCACGATTCCTCGCACACA 426
134 ysgInArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
|||||
427 AGCAGCGAAGAAGGTGATATGCAGGCTTCAGCCGCGAGGCACTCCAG 476
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGluGlnTr 167
|||||
477 TGCTACGTGCTCGTGATCGCTGAGGAAGTCAGCAGTGTCTGAGCAGTG 526
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
527 GCTAAGCTGCGCGCAGCGCGCTCTCTGCTTACCCCGAGGTGAAGCGCC 576
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
|||||

577 TCATGTCCGCATCGCCCATGCGCATCTCTGGCGCTGCGAGCCGGGTCCA 626
201 AlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGluMe 217
|||||
627 GCGGGCGGGCGGGGAGAGCAGACACAGCTCTGTGGAGGCTTTCGAGAGAT 676
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
677 GAACCGCAATCTCTCTCTCTTCCCATGTGACGTGCCCTTAGCGGCTGT 726
234 yfArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
|||||
727 ACCGGGGCGTGAAGCGCGGGAACCTTATACAGCGCGCATCGAGAGAAC 776
251 ileArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
|||||
777 ATTCGCGCCAAGATCCGCGGCTTCAAGCTACAGACCGGATGGGGGTTC 826
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
|||||
827 CAAGCAGCGCGCTGCAGCTCTGATTGAGCACCTGTGGGAGAGGGAGAGA 876
284 rgleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
877 GCGTGGATATGCAGGCACCTAAACAATCGTCAACAGAGCTCCTTTGGT 926
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
|||||
927 GGTATGAACACTACAGCCAGCTGCTGCAGACATCAGTATCACTTACCTAG 976
317 yLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLysG 334
|||||
977 ACTCTACCCACATGTCTCTCCAGAAAGTTCGAGAAGAGATAAAGACAAG 1026
334 lyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
|||||
1027 GCTTACTTTGCAAGACCAATCAAGACAACAAGTTAGACATGGAACCTTG 1076
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1077 GAACAGCTTAATACATTGGGTGTGTCTTAAGAGAGACCTGCGATTGAA 1126
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1127 TCCTCCGGTCCAGAGAGGTTTCGGGTGCTCTGAAGACTTTTGAAGCTGA 1176
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
|||||
1177 ATGATATCACGATCCCAAGGGCTGGAATGTTATTACAGTATCTGTGAC 1226
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1227 ACCCAGCATGTGCGACAGATATCTTCACTAACAGAGGAATTTAATCCCGA 1276
417 parGheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||
1277 CCGCTTATAGTGCCTCATCCAGAGGATGCTTCCCGGTTACAGCTTCAATTC 1326
434 ropheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1327 CATTTGAGAGAGGCTTCGGAGCTGTGTAGGCCAAAGATTTCAAAAAATT 1376
451 leuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
|||||
1377 CTCTTAAGATATTACAGTGGAGCTGGCTAGGCACTGTGATGGCAGCT 1426
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1427 TCTTAATGAGACCTCTCAATGAAGACAGACCCCACTGTATACCTGTGG 1476
484 spAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||
1477 ACAATCTCCCTGCAAGATTCACTACTTCCAGGGAGATATATC 1517

```
seq_name: gb_pr4:AF005418
seq_documentation_block:
LOCUS       AF005418             1743 bp      mRNA           PRI           15-DEC-1997
DEFINITION  Homo sapiens retinoic acid hydroxylase mRNA, complete cds.
ACCESSION   AF005418
VERSION     AF005418.1      GI:2688845
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1743)
AUTHORS     White,J.A., Beckett-Jones,B., Guo,Y.D., Dilworth,F.J., Bonasoro,J.,
            Jones,G. and Petkovich,M.
TITLE       cDNA cloning of human retinoic acid-metabolizing enzyme (hp450RAI)
            identifies a novel family of cytochromes P450
JOURNAL     J. Biol. Chem. 272 (30), 18538-18541 (1997)
MEDLINE     97373542
REFERENCE   2 (bases 1 to 1743)
AUTHORS     White,J.A., Beckett-Jones,B., Guo,Y., Dilworth,F.J., Bonasoro,J.,
            Jones,G. and Petkovich,M.
TITLE       Direct Submission
JOURNAL     Submitted (26-MAY-1997) Cancer Research Labs, Queen's University,
            Botterell Hall, Rm 355, Kingston, Ont K7L 3N6, Canada
FEATURES
source      Location/Qualifiers
            1..1743
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               39..1532
                  /note="hp450RAI; cytochrome P450"
                  /codon_start=1
                  /product="retinoic acid hydroxylase"
                  /protein_id="AAB8881.1"
                  /db_xref="GI:2688846"
               /translation="MGLPALLASALCTFVLPILLFLAIAIKLMDLYCVSGRDRSCALPL
               PPGTMGFPEFGETLQWVLRKQLQMRKRYGFIYKTHLFRPTVRVMGADNVRILL
               GDDRLVSVHWNPASVRTILGSGCLSNLHDSHKORKKIVIMRAFSREALCYVVPVITEEV
               GSSLEQWLSGGERGLYPEVKRLMFRIAMRILGCEPOLAGDGDSEQLVEAFEEET
               RNLFSLPIDVPFSGLYRGMKARNLIHARIEONIRAKICGLRASEAGDGDALQLEIE
               HSWERGERLDMQALKOSTELLFGHEHTTASATSLITYLGLYPHVLRKREELKSKG
               ILCKSNQDNKLDMEILEOLKYGIVIKETLRINLPVPGGFRVALKTFELNGYQIPKGM
               NVIYSICDTHDVAEIFTNKEEFNDRFMLPHBEDASRSEFIPEGGLRSCVKEEPAKI
               LLIKFTVELARHCDWQLNGRPPTMKTSPTVYVPVDNLPARPTHFHEGI"
BASE COUNT      423 a      442 c      477 g      401 t
ORIGIN
alignment_scores:
      Quality: 2441.00      Length: 497
      Ratio: 5.033      Gaps: 0
Percent Similarity: 97.586      Percent Identity: 93.360
alignment_block:
US-09-668-482-32 x AF005418 ..
Align seg 1/1 to: AF005418 from: 1 to: 1743
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
39 ATGGGGCTCCCGCGGCTGCTGGCCAGTGGCTCTGCACCTTGCTGCTGCC 88
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTIPAspLeuTYrCysValS 34
|||||
89 GGTGCTGCTCTTCTCTGCTGCGATCAAGCTCTGGGACCTGTACTGCGTGA 138
34 erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
139 GCGGGCCGCGACCGCAGTTGTGCCCTCCATTGCCCGCGGACTATGGGC 188
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
```

```
189 TTCCCTTCTTTGGGAAACCTTGCAAGATGTACTGCAGCGGAGAACTT 238
67 eLeuGlnMetLysArgArgLysTYrGlyPheIleTYrLysThrHisLeuP 84
|||||
239 CTGCAGATGAAGCGCAGGAAATACGGCTTCAATCTACAAGACGACATCTGT 288
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
289 TCGGGCGGGCCACCGCTACGGGTGATGGCGCGGACAAATGTGGCGGCATC 338
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
339 TTGCTCGAGACGACCGCGCTGGTGTCCACTGGCCAGCCTCGGTGCG 388
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
389 CACCATTTCTGGGATCTGGCTGCTCTTAACCTGCACGACTCTCGCACA 438
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
|||||
439 AGCAGCGCAAGAAAGTGATTATGCGGGCTTCAGCCGCGAGGACACTCGAA 488
151 CysTYrValLeuValIleAlaGluGluValSerSerCysLeuGluGlnTr 167
|||||
489 TGCTACGTGCGGTGATCACCAGGAAGTGCGCAGACCTGGAGCAGTG 538
167 pleuSerCysGlyGluArgGlyLeuLeuValTYrProGluValLysArgL 184
|||||
539 GCTGAGCTGCGGCGAGCGCGGCTCCTGTCTACCCGAGGTGAAGCGCC 588
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
|||||
589 TCATGTCCGAATCCCATGCGCATCTCTACTGGGCTGCGGAACCCCAACTG 638
201 AlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGlu 217
|||||
639 GCGGCGCAGCGGAGACTCCGAGCAGACAGCTGTGAGAGCCTTCGAGGAAT 688
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
689 GACCCGCAATCTCTCTCGCTGCCCATCGACAGTGCCTTCAGCGGGCTGT 738
224 yTArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
|||||
739 ACCGGGGCATGAAGCGCGGGAACCTCATTCACGCGGCATCGACAGAAAC 788
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
|||||
789 ATTGGCGCCAAGATCTGCGGGCTGCGGGCATCCGAGCGGGCCAGGGCTG 838
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTYrGluArgGlyGluA 284
|||||
839 CAAGACGCGCGCTGCAGCTGTGATCGAGCACTCGTGGAGAGAGGGAGAGC 888
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
889 GGCTGGACATGCAGGCACCTAAAGCAATCTTCAACCGAATCCTCTTTTGA 938
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTYrLeuG 317
|||||
939 GGACACGAAGACACGCGCAGTGACGACACATCTGTATCACTTAACCTGG 988
317 yLeuTYrProHisValLeuGlnLysValArgGluGluIleLysSerLysG 334
|||||
989 GCTCTACCCACATGTTCTCCAGAAAGTCGAGAGAGAGCTGAAGAGTAAG 1038
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
|||||
1039 GTTACTTTCGAAGAGCATCAAGACAACAAGTTGGACATGGAATTTTG 1088
351 GluGlnLeuLysTYrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1089 GAACAACCTTAATACATCGGGGTGTATTATTAGAGAGACCCTTCGACTGAA 1138
```

367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1139 TCCCCAGATTCCAGGAGGTTCCGGTGTCTCTGAAGACTTTGAATTAA 1188
384 snGlyTyrGlnIleProLysGlyTyrPasnValIleTyrSerIleCysAsp 400
|||||
1189 ATGATACCAAGATTCCCAAGGGCTGGAATGTATCTACAGTATCTGTGAT 1238
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1239 ACTCATGATGTGCGCAGAGATCTTCCACCAAGAGAAATTAACTCTGA 1288
417 parGpheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||
1289 CCGATTGATGCTGCTCACCACAGAGATGCATCCAGGTTCCAGCTTCATTG 1338
434 ropheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1339 CATTTGAGAGGCGCTTAGAGCTGTGTAGCCAAAGAAATTGCAAAATT 1388
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTyrGlnIle 467
|||||
1389 CTCTCAAAATATTACAGTGAGCTGGCCAGGCATTTGTGACTGGCAGCT 1438
467 uLeuAsnGlyProThrMetLysThrSerProThrValTyrProValA 484
|||||
1439 TCTAAATGACCTCTCAATGAAACCAAGTCCCAACCGTGTATCTCTGTGG 1488
484 spAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||
1489 ACAATCTCCCTGCAGATTCACCCATTTCATGGGGAATC 1529
seq_name: gb_ov:AF199462

seq_documentation_block:

LOCUS AF199462 1479 bp mRNA VRT 21-NOV-1999
DEFINITION Gallus gallus retinoic acid degrading enzyme CYP26 mRNA, complete cds.

ACCESSION AF199462
VERSION AF199462.1 GI:6456740

KEYWORDS

SOURCE chicken.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 1479)
AUTHORS Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M., Jessell,T.M. and Eichele,G.

TITLE Complementary Domains of Retinoic Acid Production and Degradation in the Early Chick Embryo

JOURNAL Dev. Biol. (2000) In press

REFERENCE 2 (bases 1 to 1479)
AUTHORS Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M., Jessell,T.M. and Eichele,G.

TITLE Direct Submission

JOURNAL Submitted (28-OCT-1999) Max Planck Institute for Experimental Endocrinology, Feodor-Lynen Strasse 7, Hannover 30625, Germany

FEATURES
source location/Qualifiers

1..1479

/organism="Gallus gallus"

/db_xref="taxon:9031"

CDS

1..1479

/note="P450RAI"

/product="retinoic acid degrading enzyme CYP26"

/protein_id="AAF09250.1"

/db_xref="GI:6456741"

/translation="MGFSALVASALCTPLPLLLFLAAVRLWLYCASGRDPSCPLPLPPTMGLPFEGETLQAVLQRRKFLQMKRYGFIYKTHLFGRPYAVMGAEENVRHILLGEHLVSVOWPASVYRTILNSHNGQHKRKVIMQAFSRDAIQHYVPVIOEEVSACLAQWLGAGPCLLVYPEVKRLMFRIMRIILGLFQPRQASPDGEQOLVEAFEEEMIRN

LFSLPIDVPESGLYRGLRARNIIRAKIEENIRAKMARKEPEGYKDALQIMEHTQGN
GEOLNMQELKESATELLFGGHETTASAATSLIAFLGLHDVLOKVRKELQKLLSGP
NOEKOLNMEFLEQLKVTGCVIKETLRISPVPVGGFRALKTELNGYQIPKGMVNIYS
ICDTHDVADLEFTDKDEFNPDREMSPSPEDSSRFSFIPGGGLRSCVCKEFAKYLKIF
TVELARSCDWQLNGPPTMTGPIYVVDNLPAKFIGFSGI"

BASE COUNT 302 a 438 c 427 g 312 t
ORIGIN

alignment_scores:

Quality: 2057.50 Length: 497
Ratio: 4.502 Gaps: 3
Percent Similarity: 91.952 Percent Identity: 80.080

alignment_block:

US-09-668-482-32 x AF199462 ..

Align seg 1/1 to: AF199462 from: 1 to: 1479

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
1 ATGGGCTTCTCCGCTCTGTCGCGCCAGCGCCCTGTGCACCTTCTGCTGCC 50
17 oLeuLeuLeuPheLeuAlaLeuLysLeuTyrPasnValTyrCysValS 34
|||||
51 CCTGCTGCTCTTCTCTGCGCGCGCTCCGCTCTGGAGCTGTACTGCGCCA 100
34 erSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
101 GCGGCGCGACCCACAGCTGTCCGCTGCCCTTGCCCCCGGCGACCATGGGG 150
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
:::|||||
151 CTGCCCTTCTTCTGCGGAGAGACGCTGCAGATGCTGTCAGCGCGGAAATT 200
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
201 CCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAAGACTCACTCT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
251 TTGGGCGGCGCCACTGTGCGGGTGATGGCGGCTGAGAATGTGCGGCACATC 300
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
301 CTGCTGGGCGACGACCGCGCTGTCTCCGTGTCAGATGGCGCCCGCTGCTGCG 350
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
351 CACCATCTTGGGCTCGGGCTGCCCTCTCCAACTTCCACAACGGGCGACACACA 400
134 ysgLnaArgLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
:::|||||
401 AACACCGCAAAAGGTGATCATGACAGGCTTCTCCCGGAGATGCCCTGCAG 450
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGluGlnTr 167
|||||
451 CACTACGTGCTGTCAATCCAAAGAGAGGTGAGTGCCTGCTGCGACAGTG 500
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
501 GCTGGGTGCCGGGCC...TGCCGTGCTGTGTATCCGAGGTGAAGCGCC 547
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
|||||
548 TCATGTTCCGCATCGCCATGCGCATCTGCTGGGCTTCCAGCGCCGCGCAG 597
201 AlaGlyGlyGlyLysPgluGlnGlnLeuValGluAlaPheGluGluMe 217
||| ::|
598 GCC...AGCCCCGATGTGAGCAGCAGCTGTGTGAGGCGCTTCGAGAGAT 644
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
| |||

645 GATCCGCAACCTCTTCTTCCTCCCATCGATGTGCCTTTCAGTGGCCTCT 694

234 YRARGGLYValLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||:::|||||||:::|||||:::|||||
695 ACCGGGCCCTGAGGCGCGCAACATCATCATGCCAAGATTGAGAGAAC 744

251 ILeArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
|||||:::||||| |||||:::|||||
745 ATCCGTGCCAAGATGGCCCCG.....AAGAGCCCGAGGGTGCTA 785

267 sLYsASPAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
|||||:::|||||:::|||||:::|||||:::
786 CAAGGATGCACCTGCAGCTGTGATGGAGCACACACAGGGTAATGGAGAGC 835

284 rGLeUASpmetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
:::|||||:::|||||:::|||||:::|||||
836 AGCTCAACATGCAGAGACTGAAGAGTCTGCCACAGAGCTGCTTTGGA 885

301 GLYHisGluThrThraLSerAlaAlaThrSerLeuIleThrTyrlEugl 317
|||||:::|||||:::|||||:::|||||:::|||||
886 GGTCACGAACAACGCGCCAGTGTCTACATCACTGATTGCCCTTCTTAG 935

317 yLeuTYrProHisValLeuGlnLysValArgGluGlnIleLysSerLysG 334
|||||:::|||||:::|||||:::|||||:::|||||
936 GCTCCACCATGATGTCTTCAGAAAGTGAGGAAGAGCTGCATTAAGG 985

334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspmetGluThrLeu 350
|||||:::|||||:::|||||:::|||||:::|||||
986 GGTTACTGTCTGTGTCACCAACAGAAAACAAATGAACATGGAGTTCTTG 1035

351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||:::|||||:::|||||:::|||||:::|||||
1036 GAGCAGCTGAAGTACACAGGCTGTGTTATCAAGAAACCCTCAGGCTGAG 1085

367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
:|||||:::|||||:::|||||:::|||||:::|||||
1086 CCGCTCTGTCTCCGTGGAGGATTCGGAATTGCACCTCAAGACCCTGAGCTGA 1135

384 snGlyTYrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysasp 400
|||||:::|||||:::|||||:::|||||:::|||||
1136 ATGGTTACCAGATCCCACAAAGGCTGGAATGTTATTACAGTATCTGCGAT 1185

401 ThrHisAspValAlaAspIlePheThrAsnLysGluLupHeAsnProas 417
|||||:::|||||:::|||||:::|||||:::|||||
1186 ACCCAGCATGTGGCCGATCTCTTACCAGAACAGAGAGTTCAACCCGGA 1235

417 pARGPheileValProHisProGluAspAlaSerArgPheSerPheileP 434
|||||:::|||||:::|||||:::|||||:::|||||
1236 CCGCTTCATGTCTCCGTCTCCGAGGACTCTCTAGGTTCA GTTCA TTC 1285

434 roPheglYgLyGlyLeuArgSerCySValGlyLysGluPheAlaLysIle 450
|||||:::|||||:::|||||:::|||||:::|||||
1286 CTTCGGTGGGGCTTGAGGAGCTCCGTAGGCAAAGAGTTGCCAAAGTC 1335

451 leuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnle 467
|||||:::|||||:::|||||:::|||||:::|||||
1336 CTTCATAAAATATTACGGTGGAGCTGGCGCGGAGCTGTGACTGGCAGCT 1385

467 uLeuAsnglyProProthrMetLysThrSerProthrValTYrProValA 484
|||||:::|||||:::|||||:::|||||:::|||||
1386 GCTAAATGGACCTCCTACATGAAGAAAAAGGGCCCATAGTGTACCTGTGG 1435

484 spAsnLeuProAlaArgPheThrTYrPheGlnGlyaspIle 497
|||||:::|||||:::|||||:::|||||:::|||||
1436 ACAATCTGCCCTGCCAAATTCATAGGTTTCAGCGGCCCAATC 1476

name: gb_htg21:AL358613

```

VERSION      AL358613.8   GI:13660841
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 166337)
AUTHORS      Plumb, B.
TITLE        Direct Submission
JOURNAL      Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
              requests: clonerequest@sanger.ac.uk
              On Apr 17, 2001 this sequence version replaced gi:13625010.
COMMENT      ----- Genome Center
              Center: Sanger Centre
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquerry@sanger.ac.uk
              ----- Project Information
              Center project name: DA348J12
              ----- Summary Statistics
              Assembly program: XGAP4; version 4.5
              Sequencing vector: plasmid; L08752; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Consensus quality: 161818 bases at least Q40
              Consensus quality: 163255 bases at least Q30
              Consensus quality: 163957 bases at least Q20
              Insert size: 165037; sum-of-contigs
              Insert size: 176845; 5.5% error; agarose-fp
              Quality coverage: 5.93x in Q20 bases; sum-of-contigs Quality
              coverage: 5.81x in Q20 bases; agarose-fp
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 14 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              1
              32410 32509: contig of 32409 bp in length
              *
              32510 32510: gap of 100 bp
              *
              52914 53013: contig of 20404 bp in length
              *
              53014 53013: gap of 100 bp
              *
              55817 55817: contig of 2804 bp in length
              *
              55818 55917: gap of 100 bp
              *
              55918 74512: contig of 18595 bp in length
              *
              74513 74612: gap of 100 bp
              *
              74613 83833: contig of 9221 bp in length
              *
              83834 83933: gap of 100 bp
              *
              83934 87001: contig of 3068 bp in length
              *
              87002 87101: gap of 100 bp
              *
              87102 97488: contig of 10387 bp in length
              *
              97489 97588: gap of 100 bp
              *
              97589 100583: contig of 2995 bp in length
              *
              100584 100683: gap of 100 bp
              *
              100684 105295: contig of 4612 bp in length
              *
              105296 105395: gap of 100 bp
              *
              105396 129702: contig of 24307 bp in length
              *
              129703 129802: gap of 100 bp
              *
              129803 132462: contig of 2660 bp in length
              *
              132463 132562: gap of 100 bp
              *
              132563 137099: contig of 4537 bp in length
              *
              137100 137199: gap of 100 bp
              *
              137200 158309: contig of 21110 bp in length
              *
              158310 158409: gap of 100 bp
              *
              158410 166337: contig of 7028 bp in length

```

seq_documentation_block:	
LOCUS	AL358613 166337 bp DNA HTG 14-APR-2001
DEFINITION	Homo sapiens chromosome 10 clone RP11-348J12, *** SEQUENCING IN
PROGRESS	***, 14 unordered pieces.
ACCESSION	AL358613

misc_feature /clone_lib="RPCI-11.2"
1..32409
/note="assembly_fragment:02593
fragment_chain:1
clone_end:T7
vector_side:left"
misc_feature 32510..52913
/note="assembly_fragment:01673
fragment_chain:1"
misc_feature 53014..55817
/note="assembly_fragment:01466
fragment_chain:2"
misc_feature 55918..74512
/note="assembly_fragment:01282
fragment_chain:2"
misc_feature 74613..83833
/note="assembly_fragment:02258
fragment_chain:2"
misc_feature 83934..87001
/note="assembly_fragment:00938
fragment_chain:2"
misc_feature 87102..97488
/note="assembly_fragment:01557
fragment_chain:2"
misc_feature 97589..100583
/note="assembly_fragment:00199
fragment_chain:2"
misc_feature 100684..105295
/note="assembly_fragment:01309"
105396..129702
/note="assembly_fragment:02028.0"
129803..132462
/note="assembly_fragment:02077"
132563..137099
/note="assembly_fragment:02005
fragment_chain:3"
137200..158309
/note="assembly_fragment:00282
fragment_chain:3"
158410..166337
/note="assembly_fragment:01837
fragment_chain:3
clone_end:SP6
vector_side:right"
BASE COUNT 46344 a 33577 c 34990 g 50115 t 1311 others
ORIGIN
alignment_scores:
Quality: 1939.50 Length: 1166
Ratio: 4.007 Gaps: 13
Percent Similarity: 41.509 Percent Identity: 39.623
alignment_block:
US-09-668-482-32 x AL358613 ..
Align seg 1/1 to: AL358613 from: 1 to: 166337
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValIleuPr 17
|||||
54914 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCTCTGCACCTTCGTGCTGCC 54963
17 OLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
|||||
54964 GCTGCTGCTCTTCTCTGGCTGCGATCAAGCTCTGGGACCTGTACTGCGTGA 55013
34 erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
55014 GCGGCCGCGACCGCAGTTGTGCCCTCCCATTTGCCCCCCCGGACTATGGGC 55063
51 pheProPhePheGlyGluThrLeuGlnMetValIleu.Gln..... 63
|||||
55064 TTCCCTTCTTTGGGAAACCTTGACAGATGTAAGTAAGGAGAGG 55113

63 63
55114 TGGGGCGGACAGGCTGCTTCCCCGAGCCCGGGGCTCTGGGCTTCT 55163
63 63
55164 GCTGAAGTCGGGCTAGCGCCCCCGGAGGATGCTATTGGCGCTAGGAG 55213
63 63
55214 CAGGGCTGGCGGAGCGCGGCGCTCCCGGCGCCCTCATGCCCACTTC 55263
64 ArgArgLysPheLeuGlnMetLysArg 72
55264 TCTCCTCCGCCCTTCTCCACACAGCGGAGAGTTCCTGCAGATGAAGCGC 55313
73 ArgLysTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVa 89
|||||
55314 AGGAATACGGCTTCATCTACAAGACCGCATCTGTCGGGCGGCCACCGT 55363
89 lArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyLysHisA 106
|||||
55364 ACGGTGATGGCGCGGACAATGTGGCGGCATCTTGCTCGGAGAGCACCC 55413
106 rLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAla 122
|||||
55414 GCGTGTGTGCTCCACTGGCCAGCGTCGTGGTCGCACCATTTCTGGGATCT 55463
123 GlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys.Lys. 138
|||||
55464 GGCTGCCCTCTTAACCTGCACGACTCTCGCAAGCAGCGCAAGAAGGT 55513
138 138
55514 GGGGGCAGAGGCGGCGCTGGACAGGAGGGGACCCCATTTATGAGCG 55563
138 138
55564 CAATTCCGGCTGATGATGCTAGCGCGGGCTAGCAGCTTGAGTGGGCT 55613
138 138
55614 AGGACCTCTGCCAGTCCAGGTTAGCTATACCAGCTCGGAGAGTGCCA 55663
138 138
55664 TGTGTCTGGCAGACTGGGGGTGTCTGGAAGGGGACGGCGTAGACGAGA 55713
139 ValI1 140
55714 GGGGCGGATGAGGCTTTTAACGCTGTCCCTCTCGGGACTCAGGTGAT 55763
140 eMetGlnAlaPheSerArgGluAlaLeuGlnCys..... 151
|||||
55764 TATGGGGGCTTCAGCCGCGAGGCACTCGAATGCTACGTGCCGGTGATCA 55813
151 151
55814 CCGANNN 55863
151 151
55864 NNN 55913
152 TyrValLeuValIleAlaGluGluValSerSerCysLeuGlnTr 167
|||||
55914 NNNNTACGTGCCGCTGATCACCGAGGAAGTGGGACAGCCTGGAGCAGT 55963
167 rPleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183
|||||
55964 GGCTGAGCTGCGGCGAGCGGCGCTCTGTGCTACCCCGAGGTGAAGCGC 56013

```
184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPr 200
|||||
56014 CTGATGTTCCGAATCCGATGCGCATCTACTGGGCTGCGAACCCTCACT 56063
200 oAlaGlyGlyGlyLeuAspGluGlnGlnLeuValGluAlaPheGluGluM 217
|||||
56064 GCGGGCGAGCGGGACTCCGAGCAGCAGCTTGTGAGGCTTCGAGGAA 56113
217 eThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
|||||
56114 TGACCCGCAATCTCTTCTCGCTGCCCATGACGTGCCCTTCAGCGGGCTG 56163
234 TyrArg..... 235
|||||
56164 TACCGGGTAAGGGCGGCAACGGGCTGCCGACTAGGGCGCGGACCTGG 56213
236 .....Gly.ValLysAlaArg 240
|||||
56214 GCGTGTGCTCACCGCGCGCGCTCTCTGCGCTCAGGGCATGAAGCGCGG 56263
241 AsnLeuIleHisAlaArgIleGluAsnIleArgAlaLysIleArgAr 257
|||||
56264 AACCTCATTCACGGCGCATCGACAGAACATTCGCGCAAGATCTGCGG 56313
257 gLeuGlnAlaThrGluProAspGlyGlyCysLysAspAlaLeuGlnLeuL 274
|||||
56314 GCTGGGGCATCCGAGCGGGCCAGGCTTGCAAAAGACGCGCTGCAGCTGT 56363
274 euIleGluHisSerTrpGluArgGlyGluArgLeuAspMetGln..... 288
|||||
56364 TGATCGAGCAGCTCGTGGAGAGGGAGAGCGGCTGGACATGCAGGTGAGT 56413
288 ..... 288
56414 AGCAGCTTCAGACAGCAGCAGGAGTTTGGTCCCTGGCTTTCCAAGC 56463
288 ..... 288
56464 GCTGTTCCTGGGGCCCCAAAGCGCGCCTGGGGCCACGTTTCTGGAG 56513
288 ..... 288
56514 TGGGGCGCGCGCTCAGACTACAGCTATGGAATCCCGAAGAAAGCTGAGA 56563
288 ..... 288
56564 CACCCGGTCAGGAGAGCTGGCGAAGGGGCTGCGGGGGAAGTGGAGCATC 56613
288 ..... 288
56614 CCCTAGCCTTTCAGAGTTTCAAAAGGAAAGTTGGAATTCGCAAAATGTT 56663
288 ..... 288
56664 AATAAAGAACCCTTGCATTTTAATAAACTAAGACTTTAACTCAGAGATT 56713
288 ..... 288
56714 TCCGGTAGAGCGGGTGCCTACTCGCCTTACTGCTCCAGCTGAACATAAGG 56763
288 ..... 288
56764 GACGTTGCATTTTGTTAAGATATTGCTTTCCTTGACTTTCGTGCAGCA 56813
288 ..... 288
56814 AAACATTTAGCCCTTCTAGTCTTCCCTCCAGAACTCTCAGTTGATTTCTG 56863
288 ..... 288
56864 AGTAATCCTTCTGTCAAAACCGCAGCAGACTTGTGAGAATGTGGTCTCA 56913
289 .....Ala...LeuLysGlnSerSerThrGluLeuLeuPheGlyG 301
```

```
|||||
56914 CTCTATTTCTTAGGCACTAAAGCAATCTTCAACCGAACTCCTCTTTGAGG 56963
301 yHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyL 318
|||||
56964 ACACGAAACACAGGCCAGTGCAGCCACATCTCTGATCATACTTACCTGGGC 57013
318 euTyrProHisValLeuGlnLysValArgGluIleLysSerLys... 333
|||||
57014 TCTACCCACATGTCTTCCAGAAGAGTGCAGAGAGAGCTGAAGAGTAAGTA 57063
333 ..... 333
57064 GGGAGACTGGGTCTGGGGGTGTCCTTATTAGCTTAGGAATTCAGCTGCT 57113
333 ..... 333
57114 CCTAGCCACTTCCGAATAAGTACAGTGTGCTGCTTCATGGAGATTTT 57163
333 ..... 333
57164 GAAAGTGCAAGGCCCAAGGCTGCGTGGGCCAGTGGCAGAATTAGCTTTG 57213
333 ..... 333
57214 TGAATAAACAGAGATGAGTCTTGAGCTGTGAGCCCACTTGGCAGGGTTT 57263
333 ..... 333
57264 GACCTTGTATTCTTACCCTCTCCACCTTTTGCTGAACCGTGAGATTC 57313
333 ..... 333
57314 TCAGATAGGTTCCACTTCTTGAATGTGTGTCCAAGGCATACATAGT 57363
333 ..... 333
57364 GTGAATAGCTGATTAGTGTGGGTGTGTGTGAGTGTGGGGTGGGGTT 57413
333 ..... 333
57414 TTTTGGGGGAACCTTGCTCTTCTTAACCTGTGTAGCAGCAGCATTCCTGGGA 57463
333 ..... 333
57464 TTGTAATAGATAGTGGATTGGGCAAGCAAAATGCCATTAGCTGCTGTT 57513
333 ..... 333
57514 TCCCCCTCCAGTCTCTCCCATCATGGGGCCCTTGGGGTTAGTCTCCAC 57563
333 ..... 333
57564 TTAAGCCCTGTTACGTCGTGGGCTGATTTTATTGGAGCACAATAATA 57613
334 .....Gly.LeuLeuCysLysSerAs 340
|||||
57614 CTGTTCACCTCTGTATGACTGTTTGTGATAGGGTTTACTTTGCAAGACAA 57663
340 nGlnAspAsnLysLeuAspMetGluThrLeuGlnLeuLysTyrIleG 357
|||||
57664 TCAAGACAACAAGTTGACATGGAATTTTGGAACTTAATATACATCG 57713
357 lYcysValIleLysGluThrLeuArgLeuAsnProProValProGlyGly 373
|||||
57714 GGTGTATTATTAAAGGAGACCTTCGACTGAATCCCCAGTTCCAGGAGGG 57763
374 pheArgValAlaLeuLysThrPheGluLeuAsn..... 384
|||||
57764 TTTCGGGTGCTGTGAAGACTTTTGAATTAATGTAAAGTTAAATTTCTCT 57813
384 ..... 384
```

Accession	Source	Location/Qualifiers	Features
57814	GenBank	1-1479	1.1479
57864	GenBank	1-1479	1.1479
57914	GenBank	1-1479	1.1479
57964	GenBank	1-1479	1.1479
58014	GenBank	1-1479	1.1479
58064	GenBank	1-1479	1.1479
58114	GenBank	1-1479	1.1479
58164	GenBank	1-1479	1.1479
58214	GenBank	1-1479	1.1479
58264	GenBank	1-1479	1.1479
58314	GenBank	1-1479	1.1479
58364	GenBank	1-1479	1.1479
58401	GenBank	1-1479	1.1479
58451	GenBank	1-1479	1.1479
58485	GenBank	1-1479	1.1479
58513	GenBank	1-1479	1.1479
58548	GenBank	1-1479	1.1479
58583	GenBank	1-1479	1.1479
58613	GenBank	1-1479	1.1479
58643	GenBank	1-1479	1.1479
58673	GenBank	1-1479	1.1479
58703	GenBank	1-1479	1.1479
58733	GenBank	1-1479	1.1479
58763	GenBank	1-1479	1.1479
58793	GenBank	1-1479	1.1479
58823	GenBank	1-1479	1.1479
58853	GenBank	1-1479	1.1479
58883	GenBank	1-1479	1.1479
58913	GenBank	1-1479	1.1479
58943	GenBank	1-1479	1.1479
58973	GenBank	1-1479	1.1479
59003	GenBank	1-1479	1.1479
59033	GenBank	1-1479	1.1479
59063	GenBank	1-1479	1.1479
59093	GenBank	1-1479	1.1479
59123	GenBank	1-1479	1.1479
59153	GenBank	1-1479	1.1479
59183	GenBank	1-1479	1.1479
59213	GenBank	1-1479	1.1479
59243	GenBank	1-1479	1.1479
59273	GenBank	1-1479	1.1479
59303	GenBank	1-1479	1.1479
59333	GenBank	1-1479	1.1479
59363	GenBank	1-1479	1.1479
59393	GenBank	1-1479	1.1479
59423	GenBank	1-1479	1.1479
59453	GenBank	1-1479	1.1479
59483	GenBank	1-1479	1.1479
59513	GenBank	1-1479	1.1479
59543	GenBank	1-1479	1.1479
59573	GenBank	1-1479	1.1479
59603	GenBank	1-1479	1.1479
59633	GenBank	1-1479	1.1479
59663	GenBank	1-1479	1.1479
59693	GenBank	1-1479	1.1479
59723	GenBank	1-1479	1.1479
59753	GenBank	1-1479	1.1479
59783	GenBank	1-1479	1.1479
59813	GenBank	1-1479	1.1479
59843	GenBank	1-1479	1.1479
59873	GenBank	1-1479	1.1479
59903	GenBank	1-1479	1.1479
59933	GenBank	1-1479	1.1479
59963	GenBank	1-1479	1.1479
59993	GenBank	1-1479	1.1479
60023	GenBank	1-1479	1.1479
60053	GenBank	1-1479	1.1479
60083	GenBank	1-1479	1.1479
60113	GenBank	1-1479	1.1479
60143	GenBank	1-1479	1.1479
60173	GenBank	1-1479	1.1479
60203	GenBank	1-1479	1.1479
60233	GenBank	1-1479	1.1479
60263	GenBank	1-1479	1.1479
60293	GenBank	1-1479	1.1479
60323	GenBank	1-1479	1.1479
60353	GenBank	1-1479	1.1479
60383	GenBank	1-1479	1.1479
60413	GenBank	1-1479	1.1479
60443	GenBank	1-1479	1.1479
60473	GenBank	1-1479	1.147

```

/organism="Xenopus laevis"
/db_xref="taxon:8355"
1. .1479
/note="RACE"
/codon_start=1
/product="retinoic acid converting enzyme"
/protein_id="AAC25158.1"
/db_xref="GI:3283172"
/translation="MDLYTLTSLACTLALPILLLITAKLME
PPGTMLPFPEGTLOMYLQRRFLQVRSQYGRITKTLFGSE
GEHKLVSVHWPASVRTLIGAGCLSNLDNEHKYTKVIAQAFS
RCSVNLWQSGPCVLYVPAIKRMFRIAMRLLLKCDPQMDRE
LESLPIDVPFSGLYRGLRANLHAQIEENKEKLEQREDEHG
EPINLQALKEASATELLFGHGTTASATSLTSLALHKDLEK
EEKELSTIEVLQQLKYTSCVIKETRLRSPVAGGFVALKTEV
IADTGEADLFPDTEKFNPDRLTPLPRDSSRFGEIPGGGVF
VVELCRNCDWELNLSFAMTSPILCPVDNLPAKFKPFSSLT"
BASE COUNT      350 a      422 c      384 g      323 t
ORIGIN
alignment_scores:
      quality: 1758.50      length: 498
      ratio: 4.080      gaps: 4
Percent Similarity: 86.546      Percent Identity: 68.072
alignment_block:
US-09-668-482-32 x AF057566 ..
Align seg 1/1 to: AF057566 from: 1 to: 1479
1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
  ||| ||| ::::::::::::::::::::::::::::::::::::|||
1 ATGATCTGTACACTCTGCTCACCAGCGCTGTGTGCACCCCTGGCACTGCC 50
17 OleuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
  ||| ||| ::::::::::::::| | | | | ::::::::::::::
51 ATTGCTGCTCTCTCTCACCGCTGCCAAGCTGGGAAGTCTACTGCTGA 100
34 eSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
  ::||| ::||| | | | | | | | | | | | | | | | | | | | |
101 GGGCAAGATGCGCGCTGCGCTAACCCACTGCCCCGGGCACCATGGGA 150
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
  ::::::::::::::| | | | | | | | | | | | | | | | | | | |
151 CTGCCCTCTTCGGAGAGACTCTGCAATGCTGTACAGAGCGCAGGTT 200
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
  ||| ||| ::::::::::| | | | | | | | | | | | | | | | | | | |
201 CCTCAAGTGAAGCGAAGCCAATACGGTGGCACTACAAAGCATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
  ||| | | | | | | | | | | | | | | | | | | | | | | | | |
251 TCGGACAGCCCCACGGTGGCGGTACAGGGCGGAGAACGTTGCACAGATC 300
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
  ||| ::| | | | | | | | | | | | | | | | | | | | | | | | |
301 CTGATGGGGGAACACAAAGCTGTGTGCTGACACTGGCGGCTCTGTGCG 350
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
  ||| ||| | | | | | | | | | | | | | | | | | | | | | | |
351 CACCATCTAGGGCGCGGTCTGTCTCCAACCTGCACGACACAGACACA 400
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
  || | | | | | | | | | | | | | | | | | | | | | | | | |
401 AGTACACCAAGAAGGTGATCGCGCAAGCCTTCTCCGGGAAGCCCTCGCT 450
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGluGlnTr 167
  ||| ||| ::| | | | | | | | | | | | | | | | | | | | | |
451 AATTACGTGCGGACAGATGAAGAAGGTGAGATGCTCCGTGAACCTGTG 500
167 pLeuSerCysGlyArgGlyLeuLeuValTyrProGluValLysArgL 184
  ||| ::| | | | | | | | | | | | | | | | | | | | | | |

```


51 CGTTTACTCTTTCGCCCGGTGAAGTGTGGAGATGTTAATGATCC 100
34 erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
101 GACGAGTCGATCCGAACTCAGAGAAGTCTCTTACCGCCAGGTACCATGGCC 150
51 pheProPhepheGlyGluThrLeuGlnMetValLeuGlnArgArgLysph 67
151 TTGCCGTTCAATTGAGAAACGCTCCAGCTGATCTCCAGAGAAGAAAGTT 200
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
201 TCTGCCGATGAACGCGCAATAACGGGTGATGAGCTGATATGTGAGCGACCTCT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
251 TCGGGAACCCGACTGTACAGGTGATGGAGCTGATATGTGAGCGACAGATT 300
101 LeuLeuGlyGluHisArgLeuValSerValHisThrProAlaSerValAr 117
301 CTGCTGGGCGAACAACAGCTGTGTCTGTTCAGTGGCCAGCATCAGTGAG 350
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
351 AACCATCTCTGGGCTCTGACACCTCTCCAAATGTCCATGAGATTCAACACA 400
134 ysgLnaArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
401 AAAACAAGAAAAAGGCCATATTAGAGGCGTTCTCTCGAGATGCTCTGGAG 450
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGluGlnTr 167
451 CACTACATTCCTCCGTGATCCAGCAGAGAGGTGAAGAGCGCCATACAGGAATG 500
167 pLeu.....SerCysGlyGluArgGlyLeuLeuValTyrProGluV 181
501 GCTGCAAAAAGACTCCTGC.....GTGCTGGTTTATCCAGAAA 538
181 allYsArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
539 TGAAGAAACTCATGTTCGATAGCTATGAGAATCCTGCTGGTTTGAA 588
198 ProGlyProAlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPh 214
589 CCAGAGCAATA.....AAGACGAGCAGCAAGCAAGAACTGTGGAGAACTTT 632
214 eglGluMetThrArgAsnLeuPheSerLeuProIleAspValProPheS 231
633 TGAGGAATGATCAAAAACCTGTCTCTTCCCAATGCAGCTTCCTTTCA 682
231 ergLLeuTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIle 247
683 GTGCTGTGTACAGGGGTTTGAGGGCAGCAATTTCATTCACATCCAAAAATT 732
248 GluGluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAs 264
733 GAGGAAAACATCAGGAAGAAATTCAGAT...GACGACAATGAAGAAACGA 779
264 pglGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluA 281
780 ACAGAAATATCAAGACGCCCTTCAGCTGTGATCGAGAACAGCAGAGAGAA 829
281 rgGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeu 297
830 GTGACGAACCTTTAGTTTGCAGGCGATGAAGAAAGACAGCTACAGACCTT 879
298 LeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleTh 314
880 CTATTTGAGGTGATGAAGAACACCGCCAGCACTGCACACCTCATTGTCA 929
314 rTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluGluIleL 331
930 GTTCTGTGGTGTGAACACAGAAAGTGTGAGAAAGTCAAGAGAGAGGTTTC 979

331 ysserLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspmet 347
980 AGGAGAAAGTTGAATGGGCATGTATACACCTGGAAAGGGCTTGAGTATG 1029
348 GluThrLeuGlnGlnLeuLysTyrIleGlyCysValIleLysGluThrLe 364
1030 GAGCTGTGGACCACTGAAAGTACACTGATGTGTGATTAAAGAGACTCT 1079
364 uArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysThrp 381
1080 TAGAATCAACCTCCTCTGTCCCGAGAGATTCAAGTCCGACACTCAAAACCT 1129
381 heGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSer 397
1130 TTGAATTGAATGGTTTACCAATTCCTAAAGATGGAACGTCATTACAGC 1179
398 IleCysAspThrHisAspValAlaAspIlePheThrAsnLysGluGluPh 414
1180 ATCTGTGACACGACGATGTGGCCGACGCTCTTCCAAACAAGAGAGATT 1229
414 easnProAspArgPheIleValProHisProGluAspAlaSerArgPheS 431
1230 CCAGCCGAGAGATTTCATGAGCAAAAGGCTTGAGAGACGGGTCCAGATTTA 1279
431 erPheIleProPheGlyGlyGlyLeuArgSerCysValGlyLysGluPhe 447
1280 ACTACATCCCTTCGAGAGAGATCCAGGATGTGTGTGGCAAAAGACTTC 1329
448 AlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAs 464
1330 GCCAAAGTGTACTCAAGATCTTTTAGTTGAGTTAAACGACGATGCAA 1379
464 pTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrValT 481
1380 TTGATTTCTCTCAACGAGACCCCGACAAATGAAGAAACAGGCCGACTATT 1429
481 yrrProValAspAsnLeuProAlaArgPheThrTyrPhe 493
1430 ACCCAGTGACAAATCTCCCTACCAATTCACACTGATTAT 1467
seq_name: gb_pr5:AF252297
seq_documentation_block: 4445 bp mRNA PRI 14-JUN-2000
LOCUS AF252297
DEFINITION Homo sapiens cytochrome P450 retinoid metabolizing protein
P450RAI-2 mRNA, complete cds.
ACCESSION AF252297
VERSION AF252297.1 GI:8515440
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS White,J.A., Ramshaw,H., Taimi,M., Stangle,W., Zhang,A.,
1 (bases 1 to 4445)
Everingham,S., Creighton,S., Tam,S.P., Jones,G. and Petkovich,M.
IDENTIFICATION of the human cytochrome P450, P450RAI-2, which is
predominantly expressed in the adult cerebellum and is responsible
for all-trans-retinoic acid metabolism
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6403-6408 (2000)
MEDLINE 20300913
REFERENCE 2 (bases 1 to 4445)
AUTHORS White,J.A., Ramshaw,H., Taimi,M., Stangle,W., Zhang,A.,
Everingham,S., Creighton,S., Tam,S.-P., Jones,G. and Petkovich,M.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2000) Cytochroma Inc., 116 Barrie Street,
Bioscience Complex, Suite 2424, Kingston, ONT K7L 3N6, Canada
FEATURES
source 1..4445
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1539
CDS

```
/codon_start=1
/product="cytochrome P450 retinoid metabolizing protein
P450RAI-2"
/db_xref="GI:8515441"
/protein_id="AAF76003.1"
/translation="MLEGLDVLVALATLAACIVSVTLILAVSQOLMQLRWATRDKS
CKLPIPKSGMGPFLIGETGHWLLQSGFQSSRREKYNVEKTHLGRPLIRVTGAENV
RKILMGEHLVSTEMPSTRMLQPNIVSNSIGDIHRNKRKVFSEKIFSHAELESYLPK
IQLVIQDTHLRAMSSHPEAINVQEAOKITFRMAIRVLLGFSIPEEDLGHLEFYQQFV
DNVSLPVDPSPSGYRGIQARQIOLQGLEKAIREKIQCTQGKDYLDALDLIESKE
HGKEMTMQELKDGTLLEIFAAYATASASTSLTMQLLKPTVLEKLRDELRAHGILHS
GGPCGEGTLRLDLSGLRYLDCAVIEVMRLFTPISGYRTVLQTFELDGFQIPKGMWSV
MYSIRDTHDTRADVFVKDYNVEDPDRFSQARSEKDKGRFHYLPFGGQVRTCLGKHLAKLF
LKVLAVELASTSRFELATRTFPRITLVPLVHPVDGLSVKFFGLDSNQNEILPETEAML
SATV"
```

BASE COUNT 893 a 1265 c 1240 g 1047 t
ORIGIN

alignment_scores:

Quality:	1048.50	Length:	495
Ratio:	2.841	Gaps:	9
Percent Similarity:	74.545	Percent Identity:	44.444

alignment_block:

US-09-668-482-32 x AAF252297 ..

Align seg 1/1 to: AAF252297 from: 1 to: 4445

```
7 LeuAlaSerAlaLeuCysThrPheValLeuProLeu.....Le 19
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
22 CTGGTGTGCGCGCTGGCCACCTCGCCGCGTGCCTGGTCCGTGACGCT 71
19 uLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValSerSera 36
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
72 GCTGCTGGCCGCTGCGCAGAGCTGTGGCAGCTGCGTGGCCGCCACTC 121
36 rgAspArgSerCysAlaLeuProLeuProProGlyThrMetGlyPhePro 52
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
122 GCGACAAGAGCTGCAAGCTGCCCATCCCAAGGATCCATGGGCTTCCCG 171
53 PhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPheLeuG1 69
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
172 CTCATCGAGAGACCGGCCACTGCGTGTGACAGGGTCTGGCTTCCAGTC 221
69 nMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuPheGlyA 86
:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
222 GTCGCGGAGGAGAGATATGGCAACGTGTCAAGACGCATTTGTTGGGGC 271
86 rgProThrValArgValMetGlyAlaAspAsnValArgArgIleLeuLeu 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
272 GGCCGCTGATACGCGTGCACCGGGCGGAGAGAACGTGGCCAAGATCCTCATG 321
103 GlyGluHisArgLeuValSerValHisTrpProAlaSerValArgThrI1 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
322 GCGGAGCACACCCTCGTAGCACCAGAGTGGCCTCGCAGCACCCGCATGTT 371
119 eLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnA 136
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
372 GCTGGGCCCCCAACACGGGTGTCCAATTCCATTGGCGACATCCACCGCAACA 421
136 rGlyLysValIleMetGlnAlaPheSerArgGluAlaLeuGlnCysTyr 152
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
422 AGCGCAAGGTCTTCTCCAAGATCTTCAGCCACGAGGCCCTGGAGAGTTAC 471
153 ValLeuValIleAlaGluGluValSerSerCysLeuGluGlnTrpLeuSe 169
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
472 CTGCCCAAGATCCAGCTGTGATCCAGACACACTGCGCGCTGGAGACAG 521
169 rcysGlyGluArgGlyLeuLeuValTyrProGluValLysArgLeuMetP 186
||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
522 CCACCCCGAG...GCCATCAACGTGTACACGAGGCCGACGAAGCTGACCT 568
```

```
186 heArgIleAlaMetArgIleLeuLeuGlyCysGlu...ProGlyProAla 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
569 TCCGCATGGCCATCCGGGTGCTGCTGGGCTTCAGCATCCCT..... 609
202 GlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluMetTh 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
610 .....GAGGAGGACCTTGGGCACCTCTTTGAGGTCTTACAGACAGTTTGT 653
218 rArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrA 235
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
654 GGACAATGCTCTTCCCTGCGCTGCACCTGCCCTTCAGTGGCTACCGGC 703
235 rgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsnIle 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
704 GGGCATTCAGGCTCGCAGATCCTGCAGAAAGGGCTGAGAGAGCCATC 753
252 ArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCysLy 268
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
754 CGGAGAGAGCTGCAGTGCACACAGGGCAAGGAC.....TACTT 791
268 sAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluArgL 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
792 GGACGCCCTGGACCTCCTCATTTGAGACAGCAAGAGGACACGGGAAGAGA 841
285 euAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGlyGly 301
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
842 TGACCATGCAGGAGCTGAAGAGACGGGACCCTGAGCTGATCTTTGGCGCC 891
302 HisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyLe 318
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
892 TATGCCACCAACGGCCAGCGCCAGACCTCACTCATCATGACACTGCTGAA 941
318 uTyrProHisValLeuGlnLysValAlaArgGluLysSerLysGlyL 335
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
942 GCACCCCACTGTGCTGGAGAGCTGCGGATGAGCTGCGGCTCATGGCA 991
335 euLeu.....CysLysSerAsnGlnAspAsnLysLeuAspMet 347
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
992 TCCTGACACAGTGGCGGCTGCCCTGTC.....GAGGGACACTGCGCCTG 1035
348 GluThrLeuGluGlnLeuLysTyrIleGlyCysValIleLysGluThrLe 364
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1036 GACACGCTCAGTGGGCTGCGCTACCTGAGCTGCGTCATCAAGAGGTGAT 1085
364 uArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysThrP 381
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1086 GCGCCTGTTACAGCCCATTTCCGGCGGCTACCGCACCTGCTGCAGACCT 1135
381 heGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSer 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1136 TCGAGCTGTATGTTTCCAGATCCCAAGGCTGAGTGTATATAGC 1185
398 IleCysAspThrHisAspValAlaAspIlePheThrAsnLysGluGluPh 414
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1186 ATCCGGGACACCCATGACACAGCGCCCGTGTTCAAAGACGTGAACGTGTT 1235
414 eAsnProAspArgPheIleValProHisProGluAspAla...SerArgP 430
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1236 CGACCCGATCGCTTACGCCAGGCGGCGGAGCGAGGACAAAGATGGCGCT 1285
430 heSerPheIleProPheGlyGlyLeuArgSerCysValGlyLysGlu 446
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1286 TCCATTACCTCCCGTTCGCTGCGGTGTCCGACCTGCTGGGCAAGCAC 1335
447 pheAlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCy 463
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1336 CTGCCCAAGCTGTTCTGAAGGTGCTGCGGCTGAGACTGGCTAGCACCAG 1385
463 sasTrpGlnLeuLeuAsnGlyPro...ProThrMetLysThrSerProT 479
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1386 CCGCTTTGAGCTGGCTACACGAGACCTTCCCGCCGATCACTTGGTCCCG 1435
479 hrValTyrProValAspAsnLeuProAlaArgPhe 490
```

1436 TCCTGCACCCCGTGATGCGCTCAGCGTCAAGTTC 1470

seq_name: gb_ba3:SYCSLLE

seq_documentation_block:
LOCUS SYCSLLE 113064 bp DNA BCT 13-FEB-1999
DEFINITION Synechocystis sp. PCC6803 complete genome, 22/27, 2755703-2868766.
ACCESSION D64003 AB001339
VERSION D64003.1 GI:1001200
KEYWORDS tRNA-fMet; 2-ketoacid dehydrogenase (malate dehydrogenase, lactate dehydrogenase); 30S ribosomal protein S18; 50S ribosomal protein L33; 6-aminohexanoate-cyclic-dimer hydrolase; 7-keto-8-aminopelargonic acid synthetase; KdtB; Mg-protoporphyrin IX monomethyl ester oxidative cyclase 66 kD subunit; MoxR protein; NADH dehydrogenase subunit 5; MoxR; PleB; acetyl-coenzyme A synthetase; acyl-CoA desaturase 1; alanine racemase; aspartate 1-decarboxylase; cyanate lyase; cytochrome P450; dTDP-glucose 4,6-dehydratase; elongation factor EF-G; endo-1,4-beta-glucanase; ferredoxin-nitrite reductase; ferredoxin-thioredoxin reductase, catalytic chain; ferrochelatase; flavoprotein; low affinity sulfate transporter; methionine aminopeptidase; methyltransferase; molybdenum cofactor biosynthesis protein A; molybdenum cofactor biosynthesis protein C; molybdopterin (MP) converting factor, subunit 2; molybdopterin biosynthesis MoeA; oligopeptide transport system permease protein; peptidyl-tRNA hydrolase; phosphoribosyl formylglycinamide cyclo-ligase; photosystem II P680 chlorophyll a apoprotein; polyA polymerase; endonuclease; replicative DNA helicase; succinate dehydrogenase iron-sulfur protein; tRNA-Cys; tRNA-fMet(exon1); tRNA-fMet(exon2); Ycf34.

SOURCE Synechocystis sp. (strain:PCC6803) DNA.

ORGANISM Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

REFERENCE 1 (bases 1 to 113064)
Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N., Sugiura,M. and Tabata,S.

TITLE

Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 648 to 928 of the genome

JOURNAL DNA Res. 2 (4), 153-166 (1995)

MEDLINE

REFERENCE

AUTHORS

2 (bases 1 to 113064)
Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hiroseawa,M., Sugiura,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K., Okumura,S., Shimo,S., Takeuchi,C., Wada,T., Watanabe,A., Yamada,M., Yasuda,M. and Tabata,S.

TITLE

Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions

JOURNAL DNA Res. 3 (3), 109-136 (1996)

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (30-AUG-1995) to the DDBJ/EMBL/GenBank databases. Satoshi Tabata, Kazusa DNA Research Institute, Gene Structure 2; 1532-3 Yanauchino, Kisarazu, Chiba 292, Japan (E-mail:tabata@kazusa.or.jp, Tel:0438-52-3933, Fax:0438-52-3934)

COMMENT

Potential protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis.

FEATURES

SOURCE

CDS

1. .113064
/organism="Synechocystis sp."
/strain="PCC6803"
/db_xref="taxon:1143"
complement(1..3960)
/note="ORF_ID:s110178"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA10440.1"
/db_xref="GI:1001201"

CDS

CDS

CDS

CDS

CDS

/translation="MSQONYNMKRFWCPRSGNISLRDGYLDDPNTKWKACNPDLVS
LEDIAEIPCLVLGEPVGKTHELEKLQRFTEETICEYSKVLGLNLRFTNLKEDLFR
DKTFSDWLEGHVHLHFLDSLDEGLNKSIAIGLVEEFKPKRSHINRLRLTCR
TLAPPSILEELKALWEESNFAIYELAPLRIDVINAVEAGFSSDDFLKAINORDIV
IDORLIVAARIAAVTIFANREAVWIDVEQGVPAEDVPQKLCFGEESNGREFEITQ
EVIQEVLDITGLFSSRGLHRMGNAHQTYAEFFAWYILQRNIDLSKIKTILYSSDSTP
KLIPOLHETAVWIASMRDYNVEQIEIKTPDVLITQTDIPTDSELRATIVANLMQYEE
KLFDRCFCNVARNYAKLKHSGLAOLKVYICDSSKKEPTRRIAIDIAIECNLSQGLD
AKLTCPSPQISIRVNAARAISSIGDQSTRKLKSLANVLPEDEQLEKAKAALMAL
PROTLAEELFQALTPSKRINLGMHRYFLESELVPQDSEDLVIALDWLKTQGRCFG
HPFEELGDAILLKAMEHVNVPGLVEGFTOVALQWKEHQKLISHSPSLQAFQESLLG
NMQRALILEENLVESIAEKDEDEDFYLTSSMTEPIISSNDVDMLEKLQDSNHEKIQ
RAMAKITIEEVEDRDTKOISEI LKATQNNISLSDVFSYFTPIELDSEEVTLRDHYL
KMKNSWKPDLDPPEKERVLDCEQLESGLSAMIWRLNREMTLLPQSKRYGHNKLD
LTQLPGKKEAEOATKRRITQGAIOYIQOQTNIDYSWIGTNTYNESELYGCKALDILT
EAPRTLDNESTDIMKRWTPILISVPSNFIDSNPRKNCRLVLSYHHSPEFLTTLLK
LIDKENQNNIFVIDRLEMCWDEKLENNLITEKVKSQKPNVERKLLKLEIGGAVQA
KDIALSTRNTYNSEFSESRQKALFAIKALITYSDPDSNVNVIWEKITADVOFGRAFE
LAASCNSGIKFDLTKQIADLYAWLVQYPHYEDPDHREVMAYIVSPRONIAFRN
NSLSQLRKGTIEACAELERLIOELPHLPWKETLTYAQONMRRETWRPFKPKLIEF
LTIPEPSNIDLSNQLSTIDKRTKMEDEPKIINNISDSTINAPVGTSGVTSNNVCCSS
PEKQRFNWEILEGVIIGIIGVATYVGIFFNGIFNDVLKDSLNGSPSKIEQAMPKN"
complement(4106..4678)
/note="ORF_ID:s110888"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA10441.1"
/db_xref="GI:1001202"
/translation="MLALQPRFKSPDEYLAYEDTDQLYELFNGELIYAVPTESGLNFQ
IATFLIKFAEIVGYRRVRGHLSEVHEPEKRNFPDLTITREEHIAQLAKRNTIRLS
MLPPLLVEVYVSPGELQRERDYIAKQOYQACGIPEYWIADPQTQTVLVLELLAGCYQ
EVGNFEKVMAIQSPTEFPQLSLTVNEIEFKDL"
4733..4930
/note="ORF_ID:ssr1513"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA10442.1"
/db_xref="GI:1673308"
/translation="MAIAKFEFPLOKSGRRLYITGFFAPKESDITDVEPLQALTVEGI
AAEYARELFAQNDNYTDCLYPK"
complement(4987..6327)
/note="ORF_ID:s110887"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA10443.1"
/db_xref="GI:1001203"
/translation="MLDIOAIADHAHQGAALGIKQYDIYSSVDDAGVEVDSGEPKQ
VEASNRAVIVRWNEKGLVGTSTTDQAGISLALETAASAFGITDNIPIDFSPE
AQPPIGGAENDVAEPAPMGDLITTLVKAEODLLASHPAITGVPIYNGLSQOTTERFYL
SAGAKRYECRSYASIVLYSRAEOEGKPRSGEMKISRDLASLDIDGLQTVKEKTL
HLDYRPIPTGKYTVFSARAFSLSLGAFSNLYNAOSILDKOSLATKETLGTAIASELL
CVSDDALHPGNISAREFDEGEGTPTRRVDILKNGVLTGLHLSAGTAKRFDTQPTGHANI
GSKITVGAHFYHYHSEKTAESYSLDQANVYIDKLQALHAGVNSLOGSFSLPFDGW
LVNKGKQSIDATVAGDFLEVLYKTIYVLEPEAEVTPGVCPLVWVEGLAITGE"
complement(6412..7251)
/note="ORF_ID:s110886"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA10444.1"
/db_xref="GI:1001204"
/translation="MULIRKYFALIMPISLNGIYLVGLLFFPSPWATESPIPLTDE
OLEVGSILTKAKAFATEAGDEVTAEKYWTLEKEPONPAVWSNRGNSRVSONKLIDEA
IADFNQAIETLAPQTDPIYLRGTALFAKGEFKAIAIDYNRVLAVNEDAFAYNNRGNA
EGGLGNWTSALIEDFOQATAIAPNFAFAQANTALALLEYIGOKTEAIOIMRLVKKYPMF
ADMRAALTAVLWVDGQOGEAESNWVAAGVDSRYQDLDWVKQIRRPDPVSISALDKFL
SLS"
7468..8031
/note="ORF_ID:s1r0895"

346 pmetglutThrleuGlulnleuLysTyrIleGlyCysValIleLysGluT 363
|||||
63041 GGCAGAAACTCTTAAAAAATGCCCTATCTAGACCAAGTGTACAGAAG 63090
363 hrleuArgleuAsnProProValProGlyGlyPheArgValAlaLeuLys 379
:|||||
63091 TATGCGGTTAATTCCTCCTGTGGCGGTGTTTCGAGAATTAAATCAA 63140
380 ThrPheGluleuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTy 396
:|||||
63141 GATGTCAATTTCAGGGGTTTTCATTTCCCAAGGCTGTTAGTTCTTA 63190
396 rSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLysGluG 413
:|||||
63191 TCAATTAGCCAAACCCACGCGCATCCTGATTTATCTCTGACCCAGAA 63240
413 IupheAsnProAspArgPheIleValProHisProGluAspAlaSerArg 429
:|||||
63241 AGTTGATCCAGAACGGTTACT.....CCTGATGGTAGTGCAACT 63281
430PheSerPheIleProPheGlyGlyLeuArgSerCy 442
|||||
63282 CATATCCTCCCTTGGCCCATGTTCCTTGGCGGTGCTTACGGGAATG 63331
442 sValGlyLysGluPheAlaLysIleLeuLysIlePheThrValGluL 459
:|||||
63332 TTTAGTAAAGAAATTGCTCGGCTGGAATGAATATTGTCACCCGAT 63381
459 euAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProPro...ThrMet 474
:|||||
63382 TAATCCACACATTTGATGTGACATGTGTACCGGGGCAAAATTAGAGTTG 63431
475 LysThrSerProThrValTyrProValAspAsnLeu 486
:|||||
63432 GTTGTACCCCTAGTCCCGGCCCAAGATATTGTG 63467
seq_name: gb_ov:AF185266
seq_documentation_block:
LOCUS AF185266 322 bp mRNA VRT 27-SEP-1999
DEFINITION Gallus gallus CYP26 (CYP26) mRNA, partial cds.
ACCESSION AF185266
VERSION AF185266.1 GI:5924309
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 322)
AUTHORS Martinez-Ceballos,E. and Burdseal,C.A.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-1999) Cell and Molecular Biology, Tulane
University, 2000 Stern Hall 6400 Freret St., New Orleans, LA 70118,
USA
FEATURES
source location/Qualifiers
1. .322
/organism="Gallus gallus"
/db_xref="taxon:9031"
<1. .>322
/gene="CYP26"
<1. .>322
/gene="CYP26"
/note="contains heme binding domain; RA-metabolizing
enzyme"
/codon_start=2
/product="CYP26"
/protein_id="AAD56546.1"
/db_xref="GI:5924310"
/translation="GFRVALKTLENGYQIPKWNVIYSICDTHDYADLFTDKDEFNP
DRFMSPSPEDSSRFSFIPFGGLRSCVGEKFAKVLKIFTVELARSCDWQLNGPPTM
KTSPT"

BASE COUNT 74 a 82 c 85 g 81 t
ORIGIN
alignment_scores:
Quality: 538.00 Length: 107
Ratio: 5.173 Gaps: 0
Percent Similarity: 97.196 Percent Identity: 90.654
alignment_block:
US-09-668-482-32 x AF185266 ..
Align seg 1/1 to: AF185266 from: 1 to: 322
373 GlyPheArgValAlaLeuLysThrPheGluleuAsnGlyTyrGlnIlePr 389
|||||
2 GGGTTTCGGTTCCTCAAGACCCCTTGAGCTGAATGTTACAGATCCC 51
389 OlySGlyTrpAsnValIleTyrSerIleCysAspThrHisAspValAla 406
|||||
52 CAAAGGCTGGAATGTTATTACAGTATCTGCGATACCCACGATGTGGCCG 101
406 spIlePheThrAsnLysGluPheAsnProAspArgPheIleValPro 422
:|||||
102 ATCTCTTACCGACAGACAGAGTTCAACCCGACCGCTTCATGCTCCG 151
423 HisProGluAspAlaSerArgPheSerPheIleProPheGlyGlyLys 439
:|||||
152 TCTCCGAGAGACTCCTCTAGTTCAGTTTCATTCCTTCGGTGGGGCTT 201
439 uArgSerCysValGlyLysGluPheAlaLysIleLeuLysIlePheT 456
:|||||
202 GAGGAGCTGCGTAGGCAAGAGTTTGCAAAAGTCCCTTAATAATATTTA 251
456 hrValGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProPro 472
:|||||
252 CGGTGAGCTGGCGGAGCTGTGACTGGCAGCTGTAAATGGACCTCCG 301
473 ThrMetLysThrSerProThr 479
|||||
302 ACAATGAAAAACAGCCCACT 322
seq_name: gb_pl2:AF318211
seq_documentation_block:
LOCUS AF318211 1494 bp mRNA PLN 15-FEB-2001
DEFINITION Taxus cuspidata 5-alpha-taxadienol-10-beta-hydroxylase mRNA,
complete cds.
ACCESSION AF318211
VERSION AF318211.1 GI:12656591
KEYWORDS
SOURCE Taxus cuspidata.
ORGANISM Taxus cuspidata
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Coniferopsida; Coniferales; Taxaceae; Taxus.
REFERENCE 1 (bases 1 to 1494)
AUTHORS Schoendorf,A., Rithner,C.D., Williams,R.M. and Croteau,R.B.
TITLE Molecular cloning of a cytochrome P450 taxane 10beta-hydroxylase
CDNA from Taxus and functional expression in yeast
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (4), 1501-1506 (2001)
PUBMED 11171980
REFERENCE 2 (bases 1 to 1494)
AUTHORS Schoendorf,A. and Croteau,R.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-2000) Institute of Biological Chemistry,
Washington State University, Pullman, WA 99164, USA
FEATURES
source location/Qualifiers
1. .1494
/organism="Taxus cuspidata"
/db_xref="taxon:99806"
1. .1494
/note="cytochrome P450-like protein"
/codon_start=1
CDS

seq_name: gb_pl1:AC006931

seq_documentation_block:
LOCUS AC006931 116205 bp DNA PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 231 of 255 of the
complete sequence. Sequence from clones F14N22, F7D19.
ACCESSION AC006931 AE002093
VERSION AC006931.5 GI:6598641
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS 1 (bases 1 to 116205)
Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
Fuji,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 116205)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Dec 17, 1999 this sequence version replaced gi:4512656.
The sequence and annotation of chromosome 2 were merged from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene
prediction programs including GRAFT
(ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
University of Washington), Genscan (Chris Burge,
http://genomic.stanford.edu/GENSCAN.html), and NetPlantGene
(http://www.cbs.dtu.dk/services/NetGene2/), searches of the
complete sequence against a peptide database and plant EST
databases at TIGR, and manual curations based on those analyses.
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by two
or more gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
identified by repeatmasker (Arian Smilt,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
and Satoshi Tabata for helpful assistance. In addition, we would
like to thank the TIGR Bioinformatics Department, especially Lixin
Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
Peterson, Michael Holmes, and Delwood Richardson for software and
database support.

This work was supported by the National Science Foundation,
Department of Energy and the US Department of Agriculture.
Address all correspondence to: at@tigr.org.

FEATURES

source
misc_feature
mRNA
gene
CDS

repeat_region
repeat_region
repeat_region
mRNA
gene
CDS

repeat_region
repeat_region
repeat_region
mRNA
gene
CDS

misc_feature
mRNA
gene
CDS

Location/Qualifiers
1. .116205
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
complement(<1. .2131)
/note="Sequence from clone F14N22"
<1226. .>2344
/gene="At2g42650"
/note="F14N22.8; predicted by genscan"
1226. .2344
/gene="At2g42650"
/note="hypothetical protein"
/protein_id="AAF18664.1"
/codon_start=1
/db_xref="GI:6598642"
/translation="MSRVSPKTVDDAVKALVKEGNEKSRTKPOLLEDEGEFYLVVAL
KKIPORNFNTNAVYRIPLPHPPLINTTEDSPCLIIDRPESGLTEEDAKKNIKSENIP
TKVVKLSKLSKSDYGSFESKRKLCDSYDMFSDRVRVIMPLPKLGKRFQSKKTPVAID
LKHMMWKEQIEKACGAAMFEMRTGSCSAIKVAKLSMESDDIVENVATLNGVDVLP
RWKYIRSLHLKLSLSLPLYTQVPIYLQKIDPLGVEEVKNGEGLAKSDVDDSSKSV
KTKKKNKGIHEVRYMDSNVSETIGDEFDSVGEDEVADDLNASGDKKKRKMSSSKS
AVSGKPDIVKSKNGQSKLKKDIDESGGGFKAKTKRR"
complement(2132. .>116205)
/note="Sequence from clone F7D19"
complement(join(<2761. .2943,3035. .3164,3258. .3372,
3450. .3489,3570. .3646,3732. .>3954))
/gene="At2g42660"
complement(<2761. .>3954)
/gene="At2g42660"
/note="F7D19.34; predicted by genscan"
complement(join(2761. .2943,3035. .3164,3258. .3372,
3450. .3489,3570. .3646,3732. .3954))
/gene="At2g42660"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD21740.1"
/db_xref="GI:4512686"
/translation="MAGRSSPGKCTWLDLNRBETVEKNEEESVDEDOITNNVRQ
YIRSNMPLRLWTPDLHLSFVAVQRLGPDRAATPKLVLEMNLKGLSIAYKSHLQMY
RSKKLEPSSRPFGAFMSGQRSYIMDISRCIPHSDLRHAYNSKTVPSRVLNQDAV
TNLGNFTLMPSSWFSGLCRNDRDSTENKTLPLEIRKKNNEKRVRDEEVSVKRLKS
MSGGIIQLPEFGNCRQKPTDDINTMLSLSTSSSEP"
3681. .3724
/rpt_family="(GA)n"
/rpt_start=1
4087. .4204
/rpt_family="(TA)n"
5756. .5806
/rpt_family="(GA)n"
complement(join(<7700. .7818,7906. .8112,8195. .8348,
8844. .8989,9294. .>9396))
/gene="At2g42670"
complement(<7700. .>9396)
/gene="At2g42670"
/note="F7D19.33; predicted by genscan"
complement(join(7700. .7818,7906. .8112,8195. .8348,
8844. .8989,9294. .9396))
/gene="At2g42670"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD21739.1"
/db_xref="GI:4512685"
/translation="MPYFAORLYNTCKASFSSSDGPTEDALEKVRNVLEKIRPSDVG
I EDAQLARSRSGPLNERNGSNOSPAPKYLHLHEDCSFSGIFCMPSPSMIPLANHPG
MTVLSKLVYGSMHVKSVDWLEPQTEPEDPSOEARPAKLVKTEMTAOSPYTLYPKS
GGNIHCFKAITHCALDILAPYSSSEHDRHCTYFRKSRREDLPGELEVDGEVTVDTWT
LLEFQPPDDFVIRIRIPYRGPIVRT"
10075. .10153
/rpt_family="(GAA)n"
join(10610. .10809,11165. .11248,11522. .11604,11693. .11994)

265 YGLY.....CysLysAspAlaLeuGlnLeuLeu.....GluHis 278
1::: ||||| ||| ||::: :::
782 AGCGAGAGAAAGAAAGATATGCTTGGCGGCTGCTTGGCGGATGATG 831
278 eTrpGluArgGlyGluArgLeuAspMetGlnAlaLeuLysGlnSer 294
::: ||| ::|||:
832 GATTTCCGATGAGAGATGTGTGACTTC.....TTG 863
295 ThrGluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSe 311
::: ||||| ::|||::||| ||::: |||
864 GTGGCTTACTGTGCGCGGTTATGAACAACCTCCACGATCATGACTCT 913
311 rLeuIleThrTyrlLeuGlyLeuTyProHisValLeuGlnLysValArg 328
::: ||||| ||| ::||| ::|||:
914 CGCCGTCAAATTCTTCACCGAGACTCCTTTAGCTCTTGCTCAACTCAAG 963
328 lGluIleLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLys 344
||||| ::::: ::::: ::::: :::
964 AAGAG.....CATGAAAAGATTAGGCAATGAAGAGATTCGTATAGT 1007
345 LeuAspMetGluThrLeuGlnLeuLysTyrlLeuGlyCysValIleLys 361
||::: ::::: ::::: :::
1008 CTTGAATGAGTGAATTACAAGTCAATGCCATTCACACATGTGTGTTAA 1057
361 sGluThrLeuArgLeuAsnProProValProGlyGlyPheArgValAla 378
:||||| ||||| ::::: ||| ||||| |||:
1058 TGAGACGCTACGAGTGGCTAACATCATCGCGGCTTTTCAGACGTGCAA 1107
378 euLysThrPheGluLeuAsnGlyTyrlGlnIleProLysGlyTrpAsnVal 394
::: ||||| ::|||::||| ||||| |||||:
1108 TGACGGATGTTGAGATCAAGGTTATATAATTCCAAAAAGGTTGAAAGTA 1157
395 IleTySerIleCysAspThrHisAspValAlaAspIlePheThrAsnLys 411
::: ||||| ::|||:
1158 TTCTCATGCTTTAGAGCGGTTTCATTAGACCCAAACCACTTCAAGATGC 1207
411 sGluGluPheAsnProAspArgPhe.....IleValProHisP 424
::: ||||| ||||| ::::: |
1208 TCCGACTTTCAACCTTGAGAGATGGCAGACCAACTCGTTAACGACAGGCC 1257
424 roGluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArg 440
||::: ::::: ||| ||||| ||||| |||
1258 CTTCATAATGTG.....TTCACACCGTTTGGTGGAGGCCAAGG 1295
441 SerCysValGlyLysGluPheAlaLysIleLeuLysIlePheThrVa 457
||| ||| ||||| ::|||::||| |||||:
1296 CTATGTCCCGTTACGAGCTGGCTTACACTCTCTGTTTCTTCA 1345
457 lGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrm 474
::|||::: ::|||
1346 CCGCCTTAGTGACAGGCTTCAGTTGG.....GTTCCCTGCAG 1380
474 eLysThrSerProThrValTyrlProValAspAsnLeuProAlaArgPhe 490
::: ::::: ::|||::: |||||:
1381 AGCAAGACAAGCTGTTTCTTCCAACTACAGAAGACGCAAGAACGGTAC 1430
491 ThrTyrlPhe 493
|||
1431 CCGATCTTC 1439

seq_name: gb_pl2:ATCYP450R

seq_documentation_block:
LOCUS ATCYP450R 1608 bp mRNA PLN 13-MAY-1996
DEFINITION A.thaliana mRNA for cytochrome P450.
ACCESSION X87367
VERSION X87367.1 GI:853718
KEYWORDS Cytochrome P450.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1608)
Szekeres,M., Nemeth,K., Koncz-Kalman,Z., Mathur,J., Kauschmann,A.,
Altman,T., Redei,G.P., Nagy,F., Scheil,J. and Koncz,C.
Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
controlling cell elongation and de-etiolation in Arabidopsis
Cell 85 (2), 171-182 (1996)
96200769
2 (bases 1 to 1608)
Szekeres,M.
Direct Submission
Submitted (22-MAY-1995) M. Szekeres, Inst. of Plant Biology,
Biological Research Center, Hungarian Academy of Sciences, PO Box
521, 6701 Szeged, HUNGARY
Location/Qualifiers
1. 1608
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone_lib="lambda gt10"
/clone="C204"
/map="14.3 cm"
48. 1466
/gene="CPD"
48. 1466
/gene="CPD"
/function="loss of gene function results in dwarf
phenotype"
/note="cytochrome P450"
/codon_start=1
/product="CYP90 protein"
/protein_id="CA60793.1"
/db_xref="GI:853719"
/db_xref="SWISS-PROT:Q42569"
/translation="MAFTAFLLLISSIAAGFLLLRLRRYRRMGLPPGSLGLPLIGET
FQLGAYKTENPEPFIDERVARYGSVFMTLFGEPFIQSADPETNRFVLQNEGLFEC
SYPASICNLGKHSLLMKGSLHKRMSLTMSFANSIIKDHMLDIDRLVRENLDWS
SRVLIMEEAKKITFELVKQLMSFDGWESESLRKEVLVIEGFFSLPLPSTTYR
KAIQARKKVAEALTVVVMKRRREEEGAEKKDMAALLAADGFSDEIYDFLVALL
VAGYETSTIMTLAVKFLTETPLALALQKEHEKIRAMKSDSYSLEWSDYKSMPTQC
VNETLVRANITIGVFRRAMTDVEIKGIKPKGWVFSSFRAVHLDPNHRKDATFNP
WRWQNSVTTGPSNVFTPEGGPRLCPGYELARVALSVFLHRLVTGFSWPAEQDKLV
FFPTTRTQKRYPIFYKRRDFAT"

BASE COUNT
ORIGIN

382 a 374 c 374 g 478 t

alignment_scores:
Quality:
Ratio:
Percent Similarity:

450.00
1.562
57.256

Length:
Gaps:
Percent Identity:

503
15
26.640

alignment_block:
US-09-668-482-32 x ATCYP450R ..

Align seg 1/1 to: ATCYP450R from: 1 to: 1608

6 LeuLeuAlaSerAlaLeuCysThrPheValLeuProLeuLeuPheLe 22
|||||::||| ::|||::||| |||
72 CTCTCTCTTCATCGCGCGGCTTCTCTCTCTACTC..... 110
22 uAlaAlaLeuLysLeuTrpAspLeuTyrlCysValSerSerArgAspArgS 39
||| ||| |||
111CGCGGTACACGTT 123
39 erCysAlaLeuProLeuProGlyThrMetGlyPheProPheGly 55
::: ||||| ||||| ::::: |||||:
124 ACCGTCGATGGTCTGCTCGCGGGAAGCCTTGTCCTCTGTGATAGA 173
56 GluThrLeuGlnMetVal.....LeuGlnArgArgLysPh 67
|||||::|||::||| ::::: |||
174 GAGACTTTTCAAGTCAATCGAGGCTTACAAACAGAGAACCTGAGCCTTT 223

[illegible]

```

361 sGIuThrLeuAArgLeuAsnProProValProGlyGlyPheAArgValAlaL 378
    ::::::::::::::: :::: ||| ||||| |||:
1058 TGAGACGCTACAGAGTGGCTAACATCATCGCGGIGTTTTCAGACGTGCAA 1107
    ::::::::::::::: :::: ||| ||||| |||:
378 eulysThrPheGluLeuAsnGlyTyrglnIleProLysGlyTrpAsnVal 394
    ::::::::::::::: :::: ||| ||||| |||:
1108 TGACGGATGTTGAGATCAAGGTTATAAAATTCCAAAGGGTGGAAGTA 1157
    ::::::::::::::: :::: ||| ||||| |||:
395 IleTySerIleCysAspThrHisAspValAlaAspIlePheThrAsnly 411
    ::::::::::::::: :::: ||| ||||| |||:
1158 TTTCATCGCTTAGAGCGGCTTCATTTAGACCCAAACCACTTCAAGATGC 1207
    ::::::::::::::: :::: ||| ||||| |||:
411 sGIuGluPheAsnProAspArgPhe.....IleValProHisP 424
    ::::::::::::::: :::: ||| ||||| |||:
1208 TCGCAGCTTCAACCCCTTGAGAGTGGCAGAGCAACTCGGTAAAGACAGGCC 1257
    ::::::::::::::: :::: ||| ||||| |||:
424 roGluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArg 440
    ::::::::::::::: :::: ||| ||||| ||||| |||:
1258 CTTCCTAATGTG.....TTCACACCGTTTGTTGGTAGAGGCCAAGG 1295
    ::::::::::::::: :::: ||| ||||| ||||| |||:
441 SerCysValGlyLysGluPheAlaAlaLysIleLeuLeuLysIlePheThrVa 457
    ::::::::::::::: :::: ||| ||||| ||||| |||:
1296 CTATGTCCCGGTTACGAGCTGGCTAGGGTTGCACCTCTCTGTTTCCTTCA 1345
    ::::::::::::::: :::: ||| ||||| ||||| |||:
457 lGIuLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrm 474
    ::::::::::::::: :::: ||| ||||| ||||| |||:
1346 CCGCCTAGTGACAGCGCTTCAGTTGG.....GTTCCCTGCAG 1380
    ::::::::::::::: :::: ||| ||||| ||||| |||:
474 eLysThrSerProThrValTyrrProValAspAsnLeuProAlaArgPhe 490
    ::::::::::::::: :::: ||| ||||| ||||| |||:
1381 AGCAAGACAAGCTGGTTTCTTCTTCCAACACTACAAGAACGCAAAACGGTAC 1430
    ::::::::::::::: :::: ||| ||||| ||||| |||:
491 ThrTyrrPhe 493
    ::|
1431 CCGATCTTC 1439

seq_name: gb_pl4:SLU54770

seq_documentation_block:
LOCUS SLU54770 1395 bp mRNA PLN 18-OCT-1996
DEFINITION Solanum lycopersicum cytochrome P450 homolog (Dwarf) mRNA, complete
          cds.
ACCESSION U54770
VERSION U54770.1 GI:1421740
KEYWORDS
SOURCE
ORGANISM tomato.
          Lycopersicon esculentum
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
          Lycopersicon.
REFERENCE
  1 (bases 1 to 1395)
  Bishop,G.J., Harrison,K. and Jones,J.D.
  The tomato Dwarf gene isolated by heterologous transposon tagging
  encodes the first member of a new cytochrome P450 family
  JOURNAL Plant Cell 8 (6), 959-969 (1996)
MEDLINE 96266705
REFERENCE 2 (bases 1 to 1395)
AUTHORS Bishop,G.
TITLE Direct Submission
JOURNAL Submitted (11-APR-1996) Gerard Bishop, Sainsbury Laboratory, John
Innes Centre, Colney Lane, Norwich, Norfolk, NR4 7UH, UK

FEATURES
          source
            1..1395
              /organism="Lycopersicon esculentum"
              /strain="GCR758"
              /db_xref="taxon:4081"
              1..1395
                /gene="dwarf"
                1..1395
                  /gene="dwarf"
                  /codon_start=1
                  /product="cytochrome P450 homolog"

```


OM of: US-09-668-482-32 to: N_Geneseq_0601:* out_format : pfs
Date: Nov 6, 2001 3:58 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09668482/runat_05112001_133354_10917/app_query.fasta.1.1680
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09668482_@CGN1_1_352 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPPX
-WAIT -THREADS=1

Search information block:
Query: US-09-668-482-32
Query length: 497
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 240.270000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	..
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09252			2612.00	4889.10	2.4e-264	1725
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12205			2612.00	4889.10	2.4e-264	1725
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09247			2437.00	4561.02	4.4e-246	1494
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12204			2437.00	4561.02	4.4e-246	1494
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60752			2425.00	4534.41	1.3e-244	2124
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12216			1853.00	3449.07	3.8e-184	4164
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09251			1734.00	3234.07	3.6e-172	1850
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12214			1734.00	3234.07	3.6e-172	1850
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12214			1251.00	2319.86	3.0e-121	2677
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09209			474.00	861.18	5.4e-40	1685
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV09209			455.50	828.44	3.6e-38	1398
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV85306			450.00	816.49	1.7e-37	1608
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV85306			450.00	816.49	1.7e-37	1608
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV85306			450.00	816.23	1.7e-37	1646
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV85306			450.00	816.21	1.7e-37	1649
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV85306			423.50	766.16	1.1e-34	1667
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV85306			406.00	733.31	7.1e-33	1648
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV85306			380.00	656.53	1.3e-28	19053
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV81874			358.50	642.31	8.3e-28	1883
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV81874			358.50	642.31	8.3e-28	1883
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV81874			350.50	627.53	5.5e-27	1835
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV81874			350.50	627.53	5.5e-27	1835
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV81874			342.00	611.24	4.5e-26	1880
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV81874			342.00	611.24	4.5e-26	1880
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAV81874			341.00	609.33	5.7e-26	1884
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAV81874			341.00	609.33	5.7e-26	1884
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV81874			337.00	602.07	1.4e-25	1838
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV81874			337.00	602.07	1.4e-25	1838
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV10237			337.00	602.07	1.4e-25	1838
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV10237			333.50	610.47	4.9e-26	491
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV10237			315.50	563.40	2.1e-23	1563
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV25214			311.50	553.90	7.0e-23	1859
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV25214			311.50	553.86	7.0e-23	1866
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV25214			311.00	554.15	6.8e-23	1673
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV60780			300.00	532.40	3.4e-21	1831
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV94656			295.00	523.53	3.4e-21	1745
/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV94656			293.00	519.26	5.9e-21	1824
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV94656			291.00	515.96	9.1e-21	1750
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV94656			291.00	515.96	9.1e-21	1750
/SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AAV87717			287.00	510.08	1.9e-20	1512
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AAV87717			287.00	510.08	1.9e-20	1512
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AAV87717			287.00	510.08	1.9e-20	1512

/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV28295	287.00	503.26	4.6e-20	2759
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV60774	285.00	505.05	3.7e-20	1691
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAV65046	282.00	499.49	7.5e-20	1676
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAV92084	282.00	499.49	7.5e-20	1676
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAV44192	282.00	499.49	7.5e-20	1676

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09252

seq_documentation_block:

ID	AAV09252	standard; cDNA; 1725 BP.
AC	AAV09252;	
DT	07-JUL-1998	(first entry)
DE	Cytochrome P450RAI isoform cDNA sequence.	
KW	Retinoid regulated gene; cytochrome P450 gene; enzyme; oxidative metabolism; P450RAI; retinoic acid; RA; promoter; ss; isoform.	
XX	Synthetic.	
XX	Key	Location/Qualifiers
FT	CDS	24..1518
FT		/*tag= a
FT		/product= "Cytochrome P450RAI isoform"
XX		
PN	WO9749832-A2.	
PD	31-DEC-1997.	
XX		
PF	23-JUN-1997;	97WO-CA00488.
XX		
PR	01-OCT-1996;	96US-0724466.
PR	21-JUN-1996;	96US-0667546.
XX		
PA	(TOOH) UNIV QUEENS KINGSTON.	
XX		
PI	Petkovich PM;	
DR	WPI; 1998-077193/07.	
DR	P-PSDB; AAW37735.	
XX		

Identifying DNA encoding inducible or suppressible cytochrome P450 -
by screening for drugs which reduce the catabolism of retinoic acid,
useful in cancer chemotherapy and the treatment of acne and
psoriasis

Disclosure; Pages 59D-59F; 113pp; English.

This nucleotide sequence encoding an isoform of cytochrome P450RAI.
Its expression is dependent on the presence of retinoic acid (RA).
The retinoid-regulated genes such as the inducible cytochrome P450RAI
gene specifically metabolises a derivative of the RA. The cytochrome
P450 gene in general produces enzymes involved in the oxidative
metabolism of endogenous and exogenous compounds. The cytochrome P450
nucleotide sequence can be used to induce or suppress the expression
of its protein. P450RAI is highly induced by RA in cell lines and
tissues. This allows for development of a drug screen using promoters
and nucleotide sequences to identify drugs which are useful for
reducing the catabolism of RA.

Sequence 1725 BP; 408 A; 425 C; 475 G; 417 T; 0 other;

alignment_scores:
Quality: 2612.00 Length: 497
Ratio: 5.256 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-668-482-32 x AAV09252 ..

Align seg 1/1 to: AAV09252 from: 1 to: 1725

```
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
  |||
25 ATGGGGCTCCCGGGCGCTGCTGGCCAGTGGCTCTGCACCTTCGTGCTGCC 74
17 OlLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
  |||
75 GCTGCTGCTCTTCTCTGGCGGCGCTCAAGCTCTGGACCTGTACTGTGTGA 124
34 erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
  |||
125 GCAGCGCGGATCGCAGCTGGCGCCCTCCCTTGCCCCCGGTACCAATGGGC 174
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
  |||
175 TTCCCATTTCTTTGGGAAACATTTGAGATGGTGTGCTTCAGCGGAGAAGTT 224
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
  |||
225 TCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAGACGCATCTGT 274
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
  |||
275 TTGGGGGGCCCCACGGGTGCGGGGTGATGGCGCGGATATGTGCGGCGCATC 324
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
  |||
325 TTGCTGGGAGAGACACCGGTTGTGTGCTGCACCTGGCCCCGCTGCTGCG 374
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
  |||
375 CACCATCTGGGCGCTGGCTGCTCTCCAACTGCACGATTCCTCGCACAA 424
134 ySGlnArgLysLysValIleMetGlnAlaPheSerArgGlnAlaLeuGln 150
  |||
425 AGCAGCGAAAGAGGTGATATTATGCAAGCCCTTCAGCCGCGAGCACTCCAG 474
151 CysTyrValLeuValIleAlaGlnGlnValSerSerCysLeuGlnGlnTr 167
  |||
475 TGCTACGTGCTCGTATCGCTGAGGAAGTCAAGCAGTGTGTCGAGCAGTGTG 524
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGlnValLysArgL 184
  |||
525 GCTAAGCTGCGCGAGCGCGGCTCTCTGCTACCCCGAGGTGAAGCGCC 574
184 euMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlnProGlyPro 200
  |||
575 TCATGTTCGCAATCGCCATGGCCATCCTGCTGGGCTGCGAGCCGGGTCCA 624
201 AlaGlyGlyGlyLysAspGlnGlnGlnLeuValGluAlaPheGluGluMe 217
  |||
625 GCGGGCGGGGGAGGAGCAGCAACAGCTCGTGGAGGCTTTTCAGAGGAGAT 674
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
  |||
675 GACCCGCAATCTTCTCTCTTCCCATTTGACGTGCCCCCTTTAGCGGCTGT 724
234 yTrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
  |||
725 ACCGGGGCGTGAAGGCGCGGAACCTTATACACGCGCGCATCGAGGAGAAC 774
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
  |||
775 ATTGCGCGCAAGATCCCGGCTTCAGGCTACAGAGCCGGATGGGGTTG 824
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
  |||
825 CAAGGACGCGCTGAGCTCTGATTGAGCACTCGTGGAGAGAGGAGAGA 874
284 rgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
  |||
```

```
875 GGCTGATATGCAAGCACCTAAACAATCGTCAACAGAGCTCCTTTGGT 924
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGl 317
  |||
925 GGTCAATGAACACTACAGCCAGTGTGCTGCACATCAGTACTACTTACCTAGG 974
317 yLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLysG 334
  |||
975 ACTTACCCACATGTCTCTCCAGAAAGTTGAGAGAAGATAAAGAGCAAGG 1024
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
  |||
1025 GCTTACTTTGCAAGAGCAATCAAGACAAAGTTAGACATGGAACCTTTTG 1074
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
  |||
1075 GAACAGCTTTAAATACATTGGGTGTGTCAATTAAAGAGACCCTGCAATTGAA 1124
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
  |||
1125 TCCTCCGGTTCCAGAGAGGTTTCGGGTGCTCTGAAGACTTTTGAGCTGA 1174
1175 ATGATATACAGATCCCAAGGCGCTGGAATGTTATTACAGTATCTGTGAC 1224
384 sngLYTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysasp 400
  |||
1175 ATGATATACAGATCCCAAGGCGCTGGAATGTTATTACAGTATCTGTGAC 1224
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
  |||
1225 ACCACGATGTGGCAGATATCTTCACTAACAAGAGGAATTTAATCCCGA 1274
417 pArgPheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
  |||
1275 CCGCTTATAGTGCCTCAVCCAGAGATGCTTCCCGGTTCAAGCTTCAATTC 1324
434 roPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
  |||
1325 CATTTGAGGAGGCGCTTCGGAGCTGTGTAGGCAAAAGATTGCAAAAATT 1374
451 leuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
  |||
1375 CTTCCTTAAGATATTTACAGTGAAGCTGGCTTAGGCACGTGATGGCAGCT 1424
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
  |||
1425 TCTAATGTGACCTCCTACAATGAAGACAAGCCCACTGTGTACCTGTGG 1474
484 spAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
  |||
1475 ACAATCTCCCTGCAAGATTCACTACTTCCAGGAGATATTC 1515
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12205
seq_documentation_block:
ID AAV12205 standard; cDNA; 1725 BP.
XX
AC AAV12205;
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolising protein mp450RAI cDNA.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ss.
XX
Mus musculus.
OS
XX
FH
FT CDS
FT Location/Qualifiers
FT 25..1518
FT /*tag= a
```

XX		
PN	WO9749815-A1.	
XX		
PD	31-DEC-1997.	
XX		
PF	23-JUN-1997;	97WO-CA00440.
XX		
PR	01-OCT-1996;	96US-0724466.
PR	21-JUN-1996;	96US-0667546.
XX		
PA	(TOOH) UNIV QUEENS KINGSTON.	
XX		
PI	Beckett BR, Jones G, Petkovich PM, White JA;	
XX		
DR	WPI; 1998-077178/07.	
DR	P-PSDB; AAW44161.	
XX		
PT	Retinoid metabolising protein - useful to develop products to treat,	
PT	e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or	
PT	ichthyosis	
XX		
PS	Claim 15; Page 62-64; 110pp; English.	

This cDNA clone codes for a novel mouse retinoid metabolising protein (see AAW44160) designated mp450RAI. The encoded protein is a retinoid oxidase that has the ability to hydroxylate retinoic acid at the 4 position of the beta-ionone ring and is inducible in epithelial cells exposed to retinoic acid. The clone was isolated from a retinoic acid-treated P19 teratocarcinoma cDNA library using human P450RAI cDNA (see AAV12204) as probe. Zebrafish, human and mouse P450RAI cDNA clones (see AAV12203-05) are claimed. They can be used in the recombinant production of P450RAI proteins. Antisense nucleic acids can be used in a claimed method for inhibiting retinoic acid hydroxylation for the treatment of cancer, inhibiting retinoic acid hydroxylation for the treatment of cancer, actinic keratosis, oral leukoplakia, secondary tumours of the head and/or neck, non-small cell lung carcinomas, basal cell carcinomas, acute promyelocytic leukaemia, skin cancer, and premalignancy associated with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter sequences (see AAV12206-08) for the P450RAI genes are also claimed.

Sequence 1725 BP; 408 A; 425 C; 475 G; 417 T; 0 other;

```
alignment_scores:
  Quality: 2612.00
  Ratio: 5.256
  Percent Similarity: 100.000
  Length: 497
  Gaps: 0
  Percent Identity: 100.000
```

alignment_block:

US-09-668-482-32 x AAV12205

Align seg 1/1 to: AAV12205 from: 1 to: 1725

1 MetGlyLeuProAlaIleuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
25 ATGGGGCTCCCGCGCTGTGGCCAGTGCCTTCACACTTCGTGCTGCC 74

17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrcysValS 34
|||||
75 GCTGCTGCTCTTCCTGGCGGCGCTCAAGCTCTGGGACCTGTACTGTGTGA 124

34 erSerArgAspargSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
125 GCAGCCGCGATCGCACGTGGCCCCCTCCCTTGCCCCCGGTACCATTGGGC 174

51 PheProPhePheglyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
175 TTCCCATTTCTTGGGGAAACAATTGCAGATGGTCTTCAGCGGAGAAGATT 224

67 eleuGlnMetLysArgArgLysTyrglyPheIleTyrlYstThrsileuP 84
|||||
225 TCTGCAGATGAAGCGCAGGAATAACGGCTTCATCTACAGACGCATCTGT 274

84	heGIARgProthrValARgValmetGLyAlaaspasValARgRIle	100
275	TTGGGGCGGCCACGCGGTGGGGTGATGGGCGCGGATATGTGGCGGCATC	324
101	LeuLeuGLyGLuHISARgLeuValSerValHIStrProAlaSerValAR	117
325	TTGCTGGGAGAGACACCGGTTGGTGTGGGTGCACTGGCCCGCGGTGGCG	374
117	gthrIleLeuGLyAlaGLyCysLeuSerAsnLeuHISpSerSerHisL	134
375	CACCATCTGGGGCGCTGGCTGCCTTCCAACCTGCACGATTCCTGCACAA	424
134	ysGLINARgLYsLYsValIleMetGLInAlaPheSerARgGLuAlaLeuGLN	150
425	AGCAGCAAAAGAAAGGTGATTATGCAGGCCCTTCAGCCGCGAGGCACCTCCAG	474
151	CysTrYrValLeuValIleAlaGLuGLuValSerSerCysLeuGLuGLINtr	167
475	TGCTACGTGCTCGTGATCGCTGAGGAGTCAGCAGTTGTGTGGAGCAGTG	524

167	plEuSerCysgIygluARgglYleuLeuValTyrProgluValLysARgl	184
525	GCTAAGCTGCGCGAGCGCGGCTCTGTGTTACCCCGAGGTGAAGCGCC	574
184	eumETphearGlleAlaMetArglleLeuLeuGlyCysgluProglYPro	200
575	TCATGTTCCGCATTCGCCATGCGCATCTGTGGCTGCGAGCCGGGTCCA	624
201	AlaGlYglYglYAspGluglInglInLeuValgluAlaPhegluGluMe	217
625	GCGGCGCGCGGAGAGACGACAACAGCTCGTGAGGCTTTCGAGAGAT	674
217	tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlYLeuT	234
675	GACCCGCAATCTCTCTCTTCCCATTGACGTGCCCCTTAAGCGGCTGT	724
234	YrArGglYValLysAlaArgAsnLeuIleHisAlaArgIlegluAsn	250
725	ACCGGGCGGTGAAGCGCGGAACCTTATACACGCGCGCATCGAGAGAAC	774
251	IleArGlAlaLysIleArgArgleuGlAlaThrgluProAspglYglYCy	267
775	ATTCGGCCCAAGATCCGCGGCTTCAGGCTACAGAGCCGATGGGGTTG	824
267	sLysAspAlaLeuGlInLeuLeuIlegluHisserTyrgluArgglYgluA	284
825	CAAGGACGCGCTGCAGCTCTCTGATTGAGCACTCGTGGAGAGGGAGAGA	874
284	rgLeuAspMetGlInAlaLeuLysGlnSerSerThrgluLeuLeuPheglY	300
875	GGCTGATATGCAGGCACTAAACAATCGTCAACAGAGCTCTCTTGGT	924
301	GlYHisgluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGl	317
925	GGTCATGAACCTACAGCCAGCTGCTGCGACATCACTGATCACTTACCTAGG	974
317	YleuTyrProHisValLeuGlInLysValARggluGlulIeLysSerLysG	334
975	ACTCTACCCACATGTCTCTCCAGAAAGTTTCGAGAAAGATTAAGACCAAGG	1024
334	lyLeuLeuCysLysSerAsnGlnAspAsnLysLysLeuAspMetgluThrLeu	350
1025	GCTTACTTTGCAAGAGCAATCAAGACAACAAGTTAGACATGGAACCTTGG	1074
351	gluInleuLeuLysTyrIleglYCysValIleLysgluThrLeuArgLeuAs	367
1075	GAACAGCCTTAATACATTGGGTGTGTCACTTAAGAGAGACCCCTGCCATTGAA	1124
367	nProProValProglYglYPheArgValAlaLeuLysThrPhegluLeuA	384
1125	TCCTCCGGGTCCAGAGAGGGTTTCGGGTGCTCTGAAGACTTTTGACGTGA	1174

384 snglyTYrglnIleProlysglyTrpasValIleTySerIleCysasp 400
|||||
1175 ATGATATACCAGATCCCAAGGCTGGAATGTTATTACAGTATCTGTGAC 1224
401 ThrHisaspValAlaaspIlePheThrAsnLysgluIupheAsnProAs 417
|||||
1225 ACCACGATGTGGCAGATATCTTCACTAACAGAGAGAAATTATCCCGA 1274
417 PArgPheIleValProHisProgluAspAlaSerArgPheSerPheIleP 434
|||||
1275 CCGCTTTATAGTGCCTCAATCCAGAGATGCTCCCGTTCACTTCATTC 1324
434 ropheglYglYglYleuArgSerCysValGlyLysgluPheAlaLysIle 450
|||||
1325 CATTTGGAGAGGCGCTCGGAGCTGTGTAGGCAAGAGTTTGCAAAATT 1374
451 leuleulysIlePheThrValgluLeuAlaArgHisCysAspTrpGlnle 467
|||||
1375 CTTCCTTAAGATATTACAGTGGAGCTGGCTAGGCACTGTGATTGGCAGCT 1424
467 uLeuAsnglyProProThrMetLysThrSerProThrValTyrrProVala 484
|||||
1425 TCTAAATGGACCTCTCTACATGAAGACAAGCCCACTGTGTACCTGTGG 1474
484 sPAsnLeuProAlaArgPheThrTyrrPheGlnIlyaspIle 497
|||||
1475 ACAATCTCCCTGCAAGATTCACTACTTCCAGGGAGATATC 1515

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09247

seq_documentation_block:

ID AAV09247 standard; cDNA; 1494 BP.

XX AAV09247;

DT 07-JUL-1998 (first entry)

XX Human cytochrome P450RAI cDNA sequence.

XX Retinoid regulated gene; cytochrome P450 gene; enzyme;

KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter; ss.

XX Homo sapiens.

OS

XX

Key Location/Qualifiers
FT CDS 1..1494
FT /*tag= a
FT /product= "Cytochrome P450RAI"

XX WO9749832-A2.

XX 31-DEC-1997.

XX 23-JUN-1997; 97WO-CA00488.

XX 01-OCT-1996; 96US-0724466.

PR 21-JUN-1996; 96US-0667546.

XX (TOOH) UNITV QUEENS KINGSTON.

XX Petkovich PM;

XX WPI; 1998-077193/07.

DR P-PSDB; AAW37734.

XX

PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis

XX

PS Example 4; Pages 56-58; 113pp; English.

XX

CC This nucleotide sequence encodes the human cytochrome P450RAI. Its

CC expression is dependent on the presence of retinoic acid (RA). The
CC retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RAI is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen
CC using promoters and nucleotide sequences to identify drugs which are
CC useful for reducing the catabolism of RA.

SQ Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other;

alignment_scores:

Quality: 2437.00 Length: 497
Ratio: 5.035 Gaps: 0
Percent Similarity: 97.384 Percent Identity: 93.360

alignment_block:

US-09-668-482-32 x AAV09247 ..

Align seg 1/1 to: AAV09247 from: 1 to: 1494

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
1 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCGCTCTGCACCTTCTGCTGCC 50
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpasPleuTyrrCysValS 34
|||||
51 GCTGCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGA 100
34 erSerArgAsparGSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
101 GCGGCCGCGACCGCAGTTGTGCCCTCCCAATGGCCCCCGGAGACTATGGGC 150
51 PheProPhePheGlygluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
151 TTCCCTCTCTTTGGGGAACCTTGACAGATGGTACTGCAGCGAGAGATT 200
67 eleuGlnMetLysArgArgLysTyrrGlyPheIleTyrrLysThrHisLeuP 84
|||||
201 CCTGCAGATGAAGCGCAGGAATACGGCTTCACTCAACAAGACGATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
251 TCGGGCGGCCACCGCTACGGGTGATGGGGCGGGAATGTGGCGGCATC 300
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
301 TTGCTCGAGACGACCGGCTGTGTGCTCCACTGGCCAGCGTGGTGG 350
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
351 CACCATTTGGGATCTGGCTGCCCTCTTAACCTGCACGACTCCTGCACA 400
134 ysgInArgLysLysValIleMetGlnAlaPheSerArggluAlaLeuGln 150
|||||
401 AGCAGCGCAGAAGGTGATTATGCGGGCTTCAAGCCGCGAGGACTTCGAA 450
151 CysTyrrValLeuValIleAlaGluGlnValSerSerCysLeuGlnInTr 167
|||||
451 TGCTACGTGCGCGGTGATCACCGAGGAAGTGGGCAAGCCTTGAGCAGTG 500
167 pleuSerCysGlygluArgGlyLeuLeuValTyrrProgluValLysArgL 184
|||||
501 GCTGAGCTGCGCGCAGCGCGGCTCTGCTTACCCCGAGGTGAAGCGCC 550
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysgluProGlyPro 200
|||||
551 TCATGTTCCGAATCGCATGCGCATCTACTGGGCTGCGAACCCTCAACTG 600
201 AlaGlyglYglYluaspGluGlnGlnLeuValgluAlaPheGluGluMe 217

601	GGCGGGCAGCGGGGACTCCGACGACAGCACTGTGTGAGGCGCTTCGAGCAAAAT	650
217	tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT	234
651	GACCCCCAATCTCTTCGCTCGCCCATCGACGTGCCCTTCAGCGGGCTGT	700
234	yTrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluIleAsn	250
701	ACCGGGGGCATGAAGGCGCGGAACCTTCATTCACGCGCGCATCGAGCAGAAC	750
251	IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCys	267
751	ATTGCGGCCAAGATCTGGCGGGCTGCGGGCATCCGAGCGGGCCAGGGCTG	800
267	sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA	284
801	CAAAAGACGCGCTGCAGCTGTGTATCGACACTCGTGGAGAGGGGAGAGC	850
284	rGluAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly	300
851	GGCTGGACATGACGCACTAAAGCAATCTTCAACCGAACTCCTCTTGA	900
301	GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG1	317
901	GGACACGAAACCAACGCGCAGTGCAGCCACATCTCTGATCACTTACCTGG	950
317	yLeuTyrProHisValLeuGlnLysValArgGluIleLysSerLysG	334
951	GCTCTACCCACATGTCTCTCCAGAAAGTCCGAGAAAGAGCTGAAGAGTAAG	1000
334	lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu	350
1001	GTTTACTTTCGCAAGACCAATCAAGACAACAAGTTGGACATGGAATTTTG	1050
351	GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs	367
1051	GAACAACCTTAATACATCGGGGTGTATTATTAAGGAGAACCTTCGACTGAA	1100
367	nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA	384
1101	TCCCCCAGTTCACAGAGGGGTTTCGGTGTCTGTGAAGACTTTTGAATTAA	1150
384	snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp	400
1151	ATGATATACCAGATTCCCAAGGGCTGGAATGTATTACTACAGTATCTGTGAT	1200
401	ThrHisAspValAlaAspIlePheThrAsnLysGluIlePheAsnProAs	417
1201	ACTCATGATGTGGCGAGAGATCTTCACCAACAAGAAAGATTAACTCTGA	1250
417	ParGpheIleValProHisProGluAspAlaSerArgPheSerPheIleP	434
1251	CCGATTCAAGTGCCTCTCACCCAGAGGATGCATCCAGGTTCACTTCATTC	1300
434	roPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle	450
1301	CATTGGAGAGGCGCTTAGAGAGCTGTGTAGCCAAAGAAATTTGCCAAAAATT	1350
451	leuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe	467
1351	CTTCTCAAAATATTACAGCTGAGAGCTGGCCAGGCATTGTGACTGGCAGCT	1400
467	uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA	484
1401	TCTAAATGGACCTCTACAAATGAAAACCAAGTCCACCGTATCTCTGTGG	1450
484	spAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle	497
1451	ACAATCTCCCTGCAGAATTCACCATTTCCATGCGGGAATC	1491

seq name: /SIDSL/qcqdqdata/qeneseq/qeneseqn/NA1998.DAT:AAV12204

```
seq_documentation_block:
ID      AAV12204 standard; cDNA; 1494 BP.
```

AAV12204; AC

DT 22-JUN-1998 (first entry)

Human retinoid metabolising protein hp45ORAI cDNA.

Retinoid metabolising protein; P450RAI; retinoid oxidase;

actinic keratosis; oral leukoplakia; head tumour; neck tumour;

KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;

XX

XX

XX
J
J
J
J
J
J
J

XX
XX
C
C
TTTTT
10067
07710-03000440

XX 01-00PM-1006, 06PM-0724466
 00

PR 21-JUN-1996; 96US-0667546.

PA (TOOH) UNIV QUEENS KINGSTON.
VV

PI Beckett BR, Jones G, Petkovich PM, White JA;
XX

DR WPI; 1998-07/11/8/01.
DR P-PSDB: AAW44160.

XX Retinoid metabolising protein - useful to develop products to treat, PT

PT	e.g. cancer
PT	ichthyosis

XX Claim 15; Page 55-57; 110pp; English.
PS

This cDNA clone codes for a novel human retinoid metabolising protein (see AAW44160) designated hp450RAI. The encoded protein is a retinoid oxidase that has the ability to hydroxylate retinoic acid at the 4 position of the beta-ionone ring and is inducible in epithelial cells exposed to retinoic acid. To isolate the clone, zebrafish P450RAI was used to search an expressed sequence tag database, and an isolated clone was used to screen a cDNA library generated from an NR2 cell line treated with retinoic acid. The hp450RAI gene has been localised to 10q23-24. zebrafish, human and mouse P450RAI cDNA clones (see AAV12203-05) are claimed. They can be used in the recombinant production of P450RAI. Antisense nucleic acids can be used in a claimed method for inhibiting retinoic acid hydroxylation for the treatment of cancer, actinic keratosis, oral leukoplakia, secondary tumours of the head and/or neck, non-small cell lung carcinomas, basal cell carcinomas, acute promyelocytic leukaemia, skin cancer, and premalignancy associated with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter sequences (see AAV12206-08) for the P450RAI genes are also claimed.

Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other;

alignment_scores:

Quality: 2437.00

Percent Similarity: 97.384 Percent Identity: 93.360

```
alignment_block:
  rc=00-6669-483-33 x AAAT12204
```

Align seg 1/1 to: AAV12204 from: 1 to: 1494

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
1 ATGGGGCTCCGGCGCTGCTGGCCAGTGGCGCTCGCACTTCGTGCTGCC 50
17 oLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
51 GCTGCTGCTCTTCCTGGCTGGATCAAGCTCTGGGACCTGTACTGCGTGA 100
34 erSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
101 GCGGCCGCGACCGCAGTGTGCCCCATTCGCCATTCGCCCGGACTATGGGC 150
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgGlyLysPh 67
151 TTCCCTCTCTTGGGGAACCTTCAGATGTGTACTGCAGCGAGGAAGTT 200
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
201 CCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAAGACGCATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
251 TCGGGCGGCCCAACCGTACGGGTGATGGCGGACAATGTGCGGCGCATC 300
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
301 TTGCTCGGAGCAGACCGGCTGTGTGGTCCACTGGCCAGCGTGGTGCG 350
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
351 CACCATCTGTGGATCTGGCTGCTCTTAACCTGCACGACTCCTCGCACA 400
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
401 AGCAGCGCAAGAAGGTGATATAGCGGCGCTTCAGCCGCGAGGCACTCGA 450
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGlnInTr 167
451 TGCTACGTGCCGGTGATCACCGAGAAGTGGGCAGCAGCCTGGAGCAGTG 500
167 pLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
501 GCTGAGCTGCGCGGAGCGCGGCTGTGTTACCCCGAGGTGAAGCGCC 550
184 euMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
551 TCATGTTCGGAATCGCCATCGGCATCTACTGGGCTGCGAACCCTCAACTG 600
201 AlaGlyGlyGlyLysAspGluGlnLeuValGluAlaPheGluGluMe 217
601 GCGGGCGACGGGGACTCCGAGCAGCAGCTTGTGAGGCGCTTCGAGGAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
651 GACCCGCAATCTTCTCGCTGCGCATCGACGTGCCCTTCAGCGGGCTGT 700
234 yrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
701 ACCGGGGCATGAAGGCGCGGAACCTCATTCACGCGCGCATCGACAGAAC 750
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
751 ATTGGCGCCAAGATCTGCGGGCTGCGGGCATCCGAGCGGCCAGGGCTG 800
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
801 CAAGAAGCGGCTGCAGCTGTGATCGAGCACTCGTGGGAGAGGGGAGAGC 850
284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
851 GGCTGGACATGCAGGCACTAAAGCAATCTTCAACCGAACTCCTTTTGA 900
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317

|||||
901 GGACACGAAGAACCGCGCGAGTGCAGCCACATCTGTGATCACTTACCTGGG 950
317 yLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLysG 334
951 GCTTACCCACATGTCTTCCAGAAAGTGGCAGAGAGCTGAAGAGTAAGG 1000
334 lYLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
1001 GTTTACTTTGCAAGACCAATCAAGACAACAAGATTGGACATGGAATTTTG 1050
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
1051 GAACAACCTTAATATCATCGGGTGTGTATTAAAGAGACCCTTCGACTGAA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
1101 TCCCCAGTTCCAGAGGGGTTTCGGGTGCTCTGAAGACCTTTGAATTAA 1150
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
1151 ATGATACAGATTCGCCAAGGGCTGGAATGTTATCTACAGTATCTGTGAT 1200
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
417 pArgPheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
1251 CCGATTCAAGTGTCTCTCACCCAGAGAGATGCATCCAGGTTCACTTCATT 1300
434 ropheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1301 CATTTGAGAGAGGCCCTTAGAGCTGTGTAGGCAAGAATTTCAAAAATT 1350
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
1351 CTCTCAAAATATTATTACAGTGGAGCTGGCCAGGCATTGTGACTGGCAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
1401 TCTAATGAGCCTCTCAATGAAGAAACCAGTCCACCGTATCTGTGTGG 1450
484 sPAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
1451 ACAATCTCCCTGCAAGATTCAACCATTTCCATGCGGAATATC 1491
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60752
seq_documentation_block:
ID AAA60752 standard; cdNA; 2124 BP.
XX
AC AAA60752;
XX
DT 31-OCT-2000 (first entry)
XX
DE Human PSEC64 encoding nucleotide sequence SEQ ID NO:1.
XX
KW Human; PSEC64; neuron growth; nerve disease; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS 46..696
FT /tag= a
FT /product= "PSEC64"
FT /note= "protein related to neuron growth"
PN JP2000152790-A.
XX
PD 06-JUN-2000.
XX
PF 19-NOV-1998; 98JP-0329989.

484 spAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||
1495 ACAATCTCCCTGCAGATTCACCCATTCCATGGGGAATC 1535

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12216

seq_documentation_block:

ID AAV12216 standard; DNA: 4164 BP.

AC AAV12216;

DT 22-JUN-1998 (first entry)

DE Mouse retinoid metabolising protein mp450RAI genomic DNA.

KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.

OS Mus musculus.

PN WO9749815-A1.

PD 31-DEC-1997.

PF 23-JUN-1997; 97WO-CA00440.

PR 01-OCT-1996; 96US-0724466.

PR 21-JUN-1996; 96US-0667546.

PA (TOOH) UNIV QUEENS KINGSTON.

PI Beckett BR, Jones G, Petkovich PM, White JA;

DR WPI; 1998-077178/07.

PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis

PS Disclosure; Page 69-70; 110pp; English.

CC This sequence comprises a genomic clone of a novel mouse retinoid
CC metabolising protein designated mp450RAI. A cDNA clone (see
CC AAV12205) that includes a coding sequence for the full-length
CC mp450RAI protein (see AAW44161) is also provided. The encoded
CC protein is a retinoid oxidase that has the ability to hydroxylate
CC retinoic acid at the 4 position of the beta-ionone ring and is
CC inducible in epithelial cells exposed to retinoic acid. Zebrafish,
CC human and mouse P450RAI cDNA clones (see AAV12203-05) are claimed.
CC They can be used in the recombinant production of P450RAI.
CC Antisense sequences can be used in a claimed method for inhibiting
CC retinoic acid hydroxylation for the treatment of cancer, actinic
CC keratosis, oral leukoplakia, secondary tumours of the head and/or
CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter
CC sequences (see AAV12206-08) for the P450RAI genes are also claimed.

CC Sequence 4164 BP; 997 A; 975 C; 1071 G; 1121 T; 0 other;

alignment_scores:

Quality: 1853.00 Length: 1046
Ratio: 4.240 Gaps: 10
Percent Similarity: 41.778 Percent Identity: 41.396

alignment_block:

US-09-668-482-32 x AAV12216 ..

Align seg 1/1 to: AAV12216 from: 1 to: 4164

40 CysAlaLeuProLeuProProGlyThrMetGlyPheProPhePheGlyI 56

73 TGGCGTCTCTCTCTCCACCTCCCTCT..... 99

56 uThrLeuGlnMetValLeuGlnArgArgLysPheLeuGlnMetLysArg 73

100CAGCGAGGAAGTTCTGCAGATGAACGCCA 130

73 rGlySTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVal 89

131 GGAATACGGCTTCATCTACAGACGCATCTGTTGGCGGCCACCGGTG 180

90 ArgValMetGlyAlaAspAsnValArgArgIleLeuGlyGluHisar 106

181 CGGGTGATGGCGCGGATATATGTGCGGCATCTGTGGAGAGCACC 230

106 gluValSerValHisTrpProAlaSerValArgThrIleLeuGlyAla 123

231 GTTGTGTGGGTGACATGGCCCGCGTGGTGGCACCACATCCTGGCGCTG 280

123 lYcysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys..Lys.. 138

281 GCTGCCTCTCCAACTGCACGATTCCTCGCACAGCAGCAAGAAGGTG 330

138 138

331 AGGGTGAGCTGGCAACTCTTGCTGGCTGGCAGGAGACCTCATCGGT 380

138 138

381 TGGTTCAGCAAAATAGAAATGCGGGCGAGGGCTAGTCCATGTGTGG 430

138 138

431 GACCAGACCCCTCTCTATCTGAGATCCACTTTAGCTTTCTGTAGCAG 480

138 138

481 TGGGTAGTCTCTGGGGGGGACTGAATTCTTGAAAGGCTACTCGGAAG 530

138 138

531 CGAAGGGGGGGGCTGAGGAAAGTAGAGATTGTAACACTCTCTGCTC 580

139ValIleMetGlnAlaPheSerArgGluAlaLeuG 150

581 CTGGGGGGTGTCTCAGGTGATTATGACGGCTTCAGCCGAGGACATCCA 630

150 nCysTrValLeuValIleAlaGluGluValSerSerCysLeuGluGln 167

631 GTGCTACGTGCCCGTGTATGCTGAGGAAGTCACAGTTGTGTGAGCAGT 680

167 rPLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183

681 GGCTAAGCTGGCGGAGCGCGGCTCTGCTTACCCGAGGTGAAGCGC 730

184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPr 200

731 CTCATGTTCGCATCGCCATGCGCATCTGCTGGGCTGCGAGCGGGTCC 780

200 oAlaGlyGlyGlyGluAspGluGlnLeuValGluAlaPheGluGluM 217

781 AGCGGGCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 830

217 etThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233

831 TGACCCGCAATCTTCTCTCTTCCATTGACGCTGCCCTTTAGCGGCTG 880

234 TyrArg..... 235

881 TACCGGTAAGGGCGTTTCGGAGTCGGAGTAGGGGAACCAAGCTCGG 930
236Gly.ValLysAlaArg 240
931 GCATCCGCTCACCGCCACGCTCTCTCCGCCCTCAGGGCGTGAGGCGCGG 980
241 AsnLeuIleHisAlaArgIleGluAsnIleArgAlaLysIleArgAr 257
|||||
981 AACCTATACACGCGCGCATCGAGAGACATTCGCGCCAAGATCCGCGG 1030
257 gLeuGlnAlaThrGluProAspGlyCysLysAspAlaLeuGlnLeuL 274
|||||
1031 GCTCAGGCTACAGAGCCGATGGGGTTGCAAGACGCGCTGCAGCTCC 1080
274 euIleGluHisSerTrpGluArgGlyLysArgLeuAspMetGln..... 288
|||||
1081 TGATTGAGCACTCGTGGAGAGGGGAGAGAGCTGATATGCAAGTGAGA 1130
288 288
1131 AGCAATTCAAAAGGTGCCAAGGGCGGGAGTGCCCTTGACTTTCCAGA 1180
288 288
1181 CACACTTTCTGGGCTCTCCAAGCCCTGTCAAGGCCCCAGCTACTTCCA 1230
288 288
1231 GTGGCGCGGATGCTAGGCTAGAGCTTTTCAACCTGTGGTCTGACCC 1280
288 288
1281 CTTACGAGCAACAACCCCTTCAGAAGGTCGCCTAAGAGCATCTGC 1330
288 288
1331 ATATCCGATATTACATCAAGAACAATACAGTAGCAAAATTACCGTTAT 1380
288 288
1381 GAAGTAGCAACAAGATAATTTATCGTTGGGGGTCAACACACAGAGG 1430
288 288
1431 AACCGTATTAAAGGTGGCATTTGGCTAGAGAGCTGTGAAGGGGTGGC 1480
288 288
1481 TGAGCAATGGGGAAGATCCAAAGTCAAAAGGCCAAGCTCATCTACAAA 1530
288 288
1531 GGTAAAGCGGAAGACAGGATTAAAGGAGTTTGGCGTTTGTGTGG 1580
288 288
1581 TCTTTGACTTTCTATGAACAACAAGGATTTTACCCTTGAAGTCTTCCGTG 1630
288 288
1631 CAATATTCTCAGGTACAGGTCTTTGTAACAGTGTATAAAGTGCACACAGA 1680
289AlaLeuLysGlnSerSerThr 295
|||||
1681 TCTGTATAAACTTCCGTTTATTATCCTTAGGCACATAAACAATCGTCAACA 1730
296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLe 312
|||||
1731 GAGCTCTCTTTGGTGTGTCATGAACATAACAGCCAGTGTCTGCAGCTCACT 1780
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluG 329
|||||
1781 GATCACTTACCTAGGACTCTACCCACATGTCTCCAGAAAGTTCGAGAAAG 1830

329 IuIleLysSer..... 332
|||||
1831 AGATAAAGACAAGGTAGATGATTCTAGAGGTTCCCATTTTGCTAGGA 1880
332 332
1881 CATTCCTCTATTAAACCACCAACCACCCCACTGTATATAAGTTGCTC 1930
332 332
1931 GATACACCCAGTACTATGACAGTGAAGATCTGAGAGCTAGTGGACTGT 1980
332 332
1981 GGGGAGAGACTCCACCCTCGTGAATTTAAAAAGCAGTTGTTTACTGG 2030
332 332
2031 GCTCTCTCTTGGGAGAAATTGACCCCTCTCCTCCTCCTCCTCCT 2080
332 332
2081 CCTCTTCCTCCTCCACCACCACCACCATCACACCCTTTATAGAGCAAGG 2130
332 332
2131 TTCTCTTTCCCTGACCAGAACAATGAATAATGTGATTAGAGCCAATAGC 2180
332 332
2181 TGATCAGGTCGACAGTGTGTGAGGGCTCAGGGTATGACCCCTTATATA 2230
332 332
2231 CCTGATAAGCAACATTTGTCTGATAATGGGTTTAGGCTGAGGAAGTGTGG 2280
332 332
2281 AAAGGAAGGCCATCAGGCCATCAGCTCTTCCCTTTATCCTCTCCCATC 2330
332 332
2331 CAGACGCCCTTCAGGTTTAGTTAACAGGTAGTCCTGCTGGGCTGACTTTT 2380
333Lys. 333
2381 TTTTGGAGTCCCCAGGATCCATCACTCTTTTATCTGTTCACATA 2430
334 .GlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrL 350
|||||
2431 GGGCTTACTTTGCAAGAGCAATCAAGACAACAAGTTAGACATGGAACCTT 2480
350 euGluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeu 366
|||
2481 TGGCACAGCTTAATAACACTGGGTGTGTCATTTAAGAGACCCCTGCGATTG 2530
367 AsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu.. 382
|||||
2531 AATCCTCCGGTTCAGAGAGGGTTTCGGGTTCCTGAAGACTTTGAGACT 2580
382 382
2581 GAATGTGAGTGCACCTCCTGTCCCCCAGCCCCAGCCCTGTCCAGTCCA 2630
382 382
2631 CTCTGTATAGCTTTGAGCATCAGCTGCCGAGAGCAGAGTGGCTACTGCC 2680
382 382
2681 TTGACAGTGTCTGCTCTCTATGTGTTACTGGGAACCAATTGCTCTCTCT 2730

```

382 ..... 382
2731 CTTAATGCCATCCATGCTAGTAATGACTTTTGTGTTGCAAGCTCAGG 2780
383 .....LeuAsn..GlyTyrG 387
2781 CCGGATGTCAATTCTTAGATTTTTTTTTTTTAAACAGGATACC 2830
387 InLeProLysGlyTrpAsnValIleYrSerIleCysAspThrHisAsp 403
2831 AGATCCCAAGGGCTGGAATGTATTACAGTATCTGTGACACCCACGAT 2880
404 ValAlaAspIlePheThrAsnLysGluGluPheAsnProAspArgPheI 420
2881 GTGGCAGATATCTTCACTAACAAGAGGAATTTAATCCCGACCGCTTTAT 2930
420 eValProHisProGluAspAlaSerArgPheSerPheIleProPheGlyG 437
2931 AGTCCTCATCCAGAGAGATGCCCGGTTCCAGCTTCATTCATTTGGAG 2980
437 IyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLeuLys 453
2981 GAGGCTTCGAGCTGTGTAGGCAAGAGTTTGCAAAAATTCCTCTTAAG 3030
454 IlePheThrValGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnG 470
3031 ATATTACAGTGAGCTGGCTAGGCACACTGTGATTGGCAGCTTCTAAATGG 3080
470 yProProThrMetLysThrSerProThrValTyrProValAspAsnLeuP 487
3081 ACCTCCTACAATGAAGACACAGCCCACTGTGTACCCCTGTGACAAATCTCC 3130
487 roAlaArgPheThrTyrPheGlnGlyAspIle 497
3131 CTGCAAGATTTACCCACTTCCAGGAGATATATC 3162
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09251

```

PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT by screening for drugs which reduce the catabolism of retinoic acid,
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis
XX
PS Disclosure: Pages 59B-59D; 113pp; English.
XX
CC This nucleotide sequence encodes the cytochrome zP450RAI of the
CC zebra fish. Its expression is dependent on the presence of retinoic
CC acid (RA). The retinoid-regulated genes such as the inducible
CC cytochrome P450RAI gene specifically metabolises a derivative of the
CC RA. The cytochrome P450 gene in general produces enzymes involved
CC in the oxidative metabolism of endogenous and exogenous compounds.
CC The cytochrome P450 nucleotide sequence can be used to induce or
CC suppress the expression of its protein. P450RAI is highly induced
CC by RA in cell lines and tissues. This allows for development of a
CC drug screen using promoters and nucleotide sequences to identify
CC drugs which are useful for reducing the catabolism of RA.
XX
SQ Sequence 1850 BP; 519 A; 391 C; 454 G; 486 T; 0 other;

```
alignment_scores:
    Quality: 1734.00      Length: 496
    Ratio:   4.070        Gaps:    4
Percent Similarity: 85.887 Percent Identity: 67.742
```

alignment_block:
US-09-668-482-32 x AAV09251 .

Align seg 1/1 to: AAV09251 from: 1 to: 1850

```

487 roAlaArpHeThrTyrPheGlnGlyAspIle 497
|||||
3131 CTGCAAGATTACCACCTTCAGGAGATATC 3162

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV09251
seq_documentation_block:
ID AAV09251 standard; cDNA; 1850 BP.
XX
AC AAV09251;
DT 07-JUL-1998 (first entry)
XX
DE Cytochrome zP450RAI cDNA sequence.
XX
KW Retinoid regulated gene; cytochrome P450 gene; enzyme;
oxidative metabolism; P450RAI; retinoic acid; RA; promoter; ss.
XX
OS Danio rerio.
XX
FH Key Location/Qualifiers
FT CDS 87..1566
FT FT /*tag= a
FT FT /product= "Cytochrome zP450RAI"
XX
PN W09749832-A2.
XX
XX 31-DEC-1997.
XX
XX 23-JUN-1997; 97WO-CA00488.
XX
XX 01-OCT-1996; 96US-0724466.
XX
XX 21-JUN-1996; 96US-0667546.
XX
XX (TOOH ) UNIV QUEENS KINGSTON.
XX
XX Petkovich PM;
XX
XX WPI; 1998-077193/07.
XX
XX P-PSDB; AAW37733.
XX

```


KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
XX Homo sapiens.
XX OS
XX PN WO9749815-A1.
XX PD 31-DEC-1997.
XX PE 23-JUN-1997; 97WO-CA00440.
XX PR 01-OCT-1996; 96US-0724466.
XX PR 21-JUN-1996; 96US-0667546.
XX PA (TOOH) UNIV QUEENS KINGSTON.
XX PI Beckett BR, Jones G, Petkovich PM, White JA;
XX DR WPI; 1998-077178/07.
XX PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
XX ichthyosis
PS Disclosure; Page 68; 110pp; English.
XX This sequence comprises a genomic clone of a novel human retinoid
CC metabolising protein designated hp450RAI. Another genomic clone
CC (see AAV12215) is also provided, as well as a cDNA clone (see AAV12204)
CC that includes a coding sequence for the full-length hp450RAI
CC protein (see AAW44160). The encoded protein is a retinoid oxidase
CC that has the ability to hydroxylate retinoic acid at the 4 position
CC of the beta-ionone ring and is inducible in epithelial cells
CC exposed to retinoic acid. The hp450RAI gene has been localised to
CC 10q23-24. Zebrafish, human and mouse P450RAI cDNA clones (see
CC AAV12203-05) are claimed. They can be used in the recombinant
CC production of P450RAI. Antisense nucleic acids can be used in a
CC claimed method for inhibiting retinoic acid hydroxylation for the
CC treatment of cancer, actinic keratosis, oral leukoplakia, secondary
CC tumours of the head and/or neck, non-small cell lung carcinomas,
CC basal cell carcinomas, acute promyelocytic leukaemia, skin cancer,
CC and premalignancy associated with actinic keratosis, acne,
CC psoriasis and/or ichthyosis. Promoter sequences (see AAV12206-08)
CC for the P450RAI genes are also claimed.
XX SQ Sequence 2677 BP; 489 A; 775 C; 868 G; 511 T; 34 other;

alignment_scores:
Quality: 1251.00 Length: 669
Ratio: 4.075 Gaps: 7
Percent Similarity: 45.889 Percent Identity: 43.797

alignment_block:
us-09-668-482-32 x AAV12214 ..

Align seg 1/1 to: AAV12214 from: 1 to: 2677

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
676 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCTGTGCACCTTCGTGCTGCC 725
17 oleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
|||||
726 GCTGCTGCTCTTCCTGGCTGGCATCAAGCTCTGGGACCTGTACTGCTGA 775
34 eSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
776 GCGGCGCGGACCGCAGTGTGGCCCTCCCATTTGCCCGGGGACTATSGGS 825
51 PheProPhePheGlyGluThrLeuGlnMetValLeu.Gln..... 63
|||||

826 TTCCCTCTCTTTGGGGAAACCTTGCAGATGNTACTNCAGGTAAGGAGGG 875
63 63
876 TGGGGCGGACAGGCTGCTTCCCCGGAGCCCCGGCGGCTCTGGGCTTCT 925
63 63
926 GCTGAAGTCGGGGTAGCGCCCGGGGAGGCATGCTATTGCGGCTAGGAG 975
63 63
976 CAGGGCTGGCGGAGCGCGGCTCCCCGGMKMCSTCAMGCSRCRWK 1025
64ArgArgLysPheLeuGlnMetLysArg 72
1026 TMMCCCTCCGCTTMCCTCCAMAGCGGARSABRNKCYKMRGATGAAGCGC 1075
73 ArgLysTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVa 89
|||||
1076 AGGAATACGCTTACTTACAAGACGCACTGTTCGGCGGCCACCGT 1125
89 lArgValMetGlyAlaAspAsnValArgArgIleLeuGlyGluHisA 106
|||||
1126 ACGGGTGATGGCGCGGACAATGTGCGGCACTGTGCTCGAGAGCACCC 1175
106 rGLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAla 122
|||||
1176 GGCTGTGTCTCGGTCCACTGGCCACGCGTCCGTCGCGCACCATTTCTGGGATCT 1225
123 GlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLys.. 138
|||||
1226 GGCTGCCCTCTTAACCTGCAGCACTCTCGCACAGCAGCGCAAGAAAGT 1275
138 138
1276 GGGGCGAGGAGGCGAGCGCTGAGACAGGAGGGGAGCCCATTTATGAGCG 1325
138 138
1326 GAATTCGGGCTGATGATGCTAGCGCGGCTAGCACCTTGAGGTGGGCT 1375
138 138
1376 AGGACCTCTGCGCAGCTCCAGGTAGCTTCCACAGCTCGAGAGTGCCAT 1425
138 138
1426 GTGTCTGGCAGGACTGGGGGTCTTGGAAGGGGAGCGCGGTAGACGAGAG 1475
139ValIle 140
1476 GGGCGGATGAGGCTTTTAACGCTGTCCCTCCTCGGAGCTCAGGTGATT 1525
141 MetGlnAlaPheSerArgGluAlaLeuGlnCysTyrValLeuValIleAl 157
|||||
1526 ATGCGGCGCTTCAAGCGGAGGCACTCGAATGCTACGTGCCGGTGATCAC 1575
157 aglGluValSerSerCysLeuGlnGlnTrpLeuSerCysGlyGluArgG 174
:|||||
1576 CGAGGAAGTGGGACAGACGCTGAGCAGTGGCTGAGCTGCGCGAGCGCG 1625
174 lYLeuLeuValTyrProGluValLysArgLeuMetPheArgIleAlaMet 190
|||||
1626 GCCTCTGTGCTACACCCGAGGTGAAGCGCTCATGTTCGAATCGCCATG 1675
191 ArgIleLeuLeuGlyCysGluProGlyProAlaGlyGlyGluAspG 207
|||||
1676 CGCATCTTACTGGGCTGCGAACCCCACTGGCGGCGAGCGGAGCTCCGA 1725
207 uGlnGlnLeuValGluAlaPheGluGlnMetThrArgAsnLeuPheSerL 224
|||||
1726 GCAGCAGCTTGTGAGGCGCTTCGAGGAAATGACCCGCAATCTTCTTCGC 1775

224 euproileaspvalpropheserglyleutytyrarg..... 235
|||||
1776 TGCCCATCGACGTGCCCTTCAGCGGGCTGTACCGGGTTAAGGCGGCAAAC 1825
235 235
1826 GGGCTGCGGACTAGGGGCGCGGACCTGGGCTGTGCTCACC GCGCG 1875
236Gly.ValLysAlaArgAsnLeuIleHisAlaArgIle 247
1876 CTCTCTGCGCTCAGGCGCATGAAGCGCGGAACCTCATTCACGCGGCATC 1925
248 GluGluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAs 264
|||:::|||||
1926 GAGCAGAACATTCGCCCAAGATCTGCGGGCTGCGGGCATCCGAGCGGG 1975
264 pGlyGlyCysLysAspAlaLeuGlnLeuIleGluHisSerTyrGlu 281
1976 CCAGGGCTGCAAGACGCGCTGCAGCTGTGATCGAGCACTCGTGGGAGA 2025
281 rgGlyLysArgLeuAspMetGln..... 288
|||||
2026 GGGAGAGCGGCTGGACATGCAGGTAGTACAGCTTCAGACCAGGCACT 2075
288 288
2076 GCGGAGTTTGGTCCCCCTGGCTTCCAAGCGCTGTCTCTGGGCCCCCAA 2125
288 288
2126 AGCGCGGCTGGGGCCAGCTTCTGGAAGTGGGGCGCGGCTCAGACTA 2175
288 288
2176 CAGCTATGGAATCCCGAAGAGGCTGAGACACCCGGTCAGAGAGCTGC 2225
288 288
2226 GGAAGGGCTGCGGGMGAACCTGGAGCATCCCTAGCCTTTAMCAGGTT 2275
288 288
2276 TCAAAGGGAAGTTGGAATTTGCAAAATGTTAATAAGAACCTTGCAT 2325
288 288
2326 TTTAATAAACTAAGACTTTAAGTCTGAGAGTTCCGCTAGCGGGGCTCG 2375
288 288
2376 TACTCGCCTTACTGCTCCAGCTGAACTAAAGGAGCTTCATTTTGTTA 2425
288 288
2426 AAGATATTGCTTCTTCTGACTTCTGTACGAAACATTTAGCCCTTCTA 2475
288 288
2476 GTCTCCCTCCAGAACTCTCAGTTCGATTCTGAGTAATCTTCTGTCAA 2525
289Ala..Leu 290
2526 CCGCAGGCGAGACTTGTGAGAAATGGGTCTCAGTCTAATCTTAGGCACTA 2575
291 LysGlnSerSerThrGluLeuLeuPheGlyGlnIsglutThrThrAlase 307
|||||
2576 AAGCAATCTTCAACCGAACTCCTTTGGAGACACGAAACACGCGCAG 2625
307 rAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuG 324
|||||
2626 TGCAGCCACATCTGTGATCACTTACCTGGGGCTCTACCCACATGTCTCC 2675

324 ln 324
2676 AG 2677
seq_name: /SIDSl/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC39029
seq_documentation_block:
ID AAC39029 standard; DNA; 1685 BP.
XX AAC39029;
AC
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 23101.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147202.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.

PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149930.
PR	23-AUG-1999;	99US-0150566.
PR	25-AUG-1999;	99US-0150884.
PR	26-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

alignment_scores:	
Quality:	474.00
Ratio:	1.612
Percent Similarity:	59.878
	Length: 491
	Gaps: 12
Percent Identity:	27.291

alignment_block:

Align seg 1/1 to: AAC39029 from: 1 to: 1685

15 ValLeuProLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTy 31

142 GTGTACCAATCCTCATCTTATGCTTGTGTA.....AGACT 182
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

964 ACGACAGCTAGTGTGCTTAACCTTGATCTTAACTTA..... 1002
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
227

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

[illegible]

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC48157

seq_documentation_block:

ID	standard; DNA; 1646 BP.
AAC48157	

AC AAC48157;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 56467.

KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS *Arabidopsis thaliana*.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

09-MAR-1999; 99US-0123548.

25-MAR-1999; 99US-0126264.

01-APR-1999; 99US-01.27462.

08-APR-1999; 99US-0128714.

19-APR-1999; 99US-0130077.

23-APR-1999; 99US-0130510.

28-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132407.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132488.

11-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134370.
10-MAY 1999 0070 0134370

19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135134

21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135353.

25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136022

28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137333

03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137503

07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138004

10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847

14-JUN-1999; 9905-0139119.
16-JUN-1999; 9905-0139452

16-JUN-1999; 9905-0139453.
17-JUN-1999; 9905-0139492

18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455

18-JUN-1999:	99US-0139457
18-JUN-1999:	99US-0139456.

18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459

18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461

18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.

18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.

22-JUN-1999: 99US-0139899.

3305-0140333. 15551 NOV 67

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality:	450.00	Length:	503
Ratio:	1.562	Gaps:	15
Percent Similarity:	57.256	Percent Identity:	26.640

alignment_block:

us-09-668-482-32 x AAC48157 ..

Align seg 1/1 to: AAC48157 from: 1 to: 1646

6 LeuLeuAlaSerAlaLeuCysThrPheValLeuProLeuLeuPheLe 22
|||||:::||||::: |||:::|||| |||
83 CTCCTCTCTTCATCGCCGCCGCGCTTCCTCCTCCTACTC..... 121
22 uAlaAlaLeuLysLeuTrpAspLeuTyrCysValSerSerArgAspArg 39
122CGCCGTACACGTT 134
39 erCysAlaLeuProLeuProProGlyThrMetGlyPheProPhePheGly 55
::: |||||||||||:::|||||:::|||||
135 ACCGTCGATGGCGTCCTCCCGGAGACCTTGCTCCCTGATAGA 184
56 GluThrLeuGlnMetVal.....LeuGlnArgArgLysPh 67
|||||:::|||||::: |||::: |||
185 GAGACTTTTCAGCTGATCGGAGCTTACAAACAGAGAACCCCTGACCTTT 234

```

67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
   ::::: ::::: ::::: ::::: ::::: :::::
235 CATGACGAGAGAGTAGCCCGGTACGGTTGGTTTCATGACGCATCTT 284
   ||||| ::::: ::::: ::::: :::::
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
   ||||| ::::: ::::: ::::: :::::
285 TTGGTGAACCGACGATTTTCTCAGCTGACCCGGAACGACCGGTTTGT 334
   ||||| ::::: ::::: ::::: :::::
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
   ||||| ::::: ::::: ::::: :::::
335 CTCAGAACGAGGAGGAGCTTTTGTAGTGTCTTATCCTGCTTCCATTTG 384
   ||||| ::::: ::::: ::::: :::::
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerHisL 134
   ::::: ||||| ::::: ::::: :::::
385 TAACCTTTGGGGAACACTCTCTGCTTCTTATGAAGGTTCTTTGCATA 434
   ::::: ||||| ::::: ::::: :::::
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgGlnAlaLeuGln 150
   ||||| ::::: ::::: ::::: :::::
435 AACGTATGCACCTCTCTCACCATGAGCTTGTCTAATCTTCAATCATTA 484
   ||||| ::::: ::::: ::::: :::::
151 CysTyrValLeuVal..IleAlaGlnValSerSerCysLeuGln 166
   ::::: ||||| ::::: ::::: :::::
485 GACCATCTCATGCTTGATATTTGACCGGTTAGTCCGGTTAATCTTGATTC 534
   ||||| ::::: ::::: ::::: :::::
166 nTriPLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysA 183
   ||||| ::::: ::::: ::::: :::::
535 TTGGCTCTCT.....CGTGTCTCTCTCATGGAAGAACCAAAA 572
   ||||| ::::: ::::: ::::: :::::
183 rGleuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGly 199
   ::::: ||||| ::::: ::::: :::::
573 AGATAACGTTTGAGCTAACGGGTGAAGCAGTTGATGAGCTTTGATCCA... 619
   ||||| ::::: ::::: ::::: :::::
200 ProAlaGlyGlyGluAspGluGlnGlnLeuValGlnAlaPheGlu 216
   ||||| ::::: ::::: ::::: :::::
620 .....GGGGAATGAGAGTGAGAGTTTAAGGAAGAGTATCTTCT 657
   ||||| ::::: ::::: ::::: :::::
216 uMetThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyL 233
   ::::: ||||| ::::: ::::: :::::
658 TGTCAATCGAAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 704
   ||||| ::::: ::::: ::::: :::::
233 euTyrArg..GlyValLysAlaArgAsnLeuIleHisAlaArgIleGlu 248
   ||||| ::::: ::::: ::::: :::::
705 CTTACCGCAAGCCATCCACGCGGAGG.....AAGGTGGCG 742
   ||||| ::::: ::::: ::::: :::::
249 GluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspG 265
   ||||| ::::: ::::: ::::: :::::
743 GAGCGGTTGACGGTGTGGTGATGAAAAAGAGGAGGAGGAGGAAGAGG 792
   ||||| ::::: ::::: ::::: :::::
265 yGly.....CysLysAspAlaLeuGlnLeuIle.....GluHis 278
   ||||| ::::: ::::: ::::: :::::
793 AGCGGAGAGAAAGATATGCTTGGCGGCTTGGCGGAGATGATG 842
   ||||| ::::: ::::: ::::: :::::
278 eTrpGluArgGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSer 294
   ||||| ::::: ::::: ::::: :::::
843 GATTTTCCGATGAAGAGATTGTGACTTC.....TTG 874
   ||||| ::::: ::::: ::::: :::::
295 ThrGluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaIleThrSe 311
   ::::: ||||| ::::: ::::: :::::
875 GTGCTTACTTGTGCGCGGTATGAAACAACTCCACGATCATGACTCT 924
   ||||| ::::: ::::: ::::: :::::
311 rLeuIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArg 328
   ::::: ||||| ::::: ::::: :::::
925 CGCCGTCAAATTTCTCACCGAGACTCTTACCTTGTCTCAACTCAAGG 974
   ||||| ::::: ::::: ::::: :::::
328 LuGluIleLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLys 344
   ||||| ::::: ::::: ::::: :::::
975 AAGAG.....CATGAAGAAGATTAGGCAATGAAGATGATTCGTATAGT 1018
   ||||| ::::: ::::: ::::: :::::
345 LeuAspMetGluThrLeuGlnLeuLysTyrIleGlyCysValIleLys 361
   ||||| ::::: ::::: ::::: :::::
1019 CTTGAATGAGATGATTAAGATCAATGCCATTCACACATGTGTGGTTAA 1068
   ||||| ::::: ::::: ::::: :::::
361 sGluThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaL 378

```

```

1069 TGAGACGCTACGAGTGGCTAACATCATCGCGCGGTGTTTCAGACGTGCAA 1118
   ::::: ||||| ::::: |||||
378 euLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnVal 394
   ::::: ||||| ::::: |||||
1119 TGACGGATGTTGAGATCAAGGTTATATAATTCGAAAGGGTGGAAGTA 1168
   ::::: ||||| ::::: |||||
395 IleTyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLys 411
   ::::: ||||| ::::: |||||
1169 TTCATATCGTTTGAAGCGGTTCATTTAGACCCAAACCACTTCMAAGATGC 1218
   ::::: ||||| ::::: |||||
411 sGluGluPheAsnProAspArgPhe.....IleValProHisP 424
   ::::: ||||| ::::: |||||
1219 TCGCACTTTCAACCTTGGAGATGGCAGAGCAACTCGGTTAACGACAGGCC 1268
   ::::: ||||| ::::: |||||
424 roGluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArg 440
   ||||| ::::: |||||
1269 CTTCTAATGTG.....TTCAACACCGTTTGGTGAGGCCCAAGG 1306
   ||||| ::::: |||||
441 SerCysValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrVa 457
   ||||| ::::: |||||
1307 CTATGTCCCGGTTACGAGCTGGCTAGGGTTGCACTCTCTGTTTCCCTTCA 1356
   ||||| ::::: |||||
457 lGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrm 474
   ::::: ||||| ::::: |||||
1357 CGCCCTAGTGACAGAGGCTTCAAGTTGG.....GTTCTGTGAG 1391
   ||||| ::::: |||||
474 euLysThrSerProThrValTyrProValAspAsnLeuProAlaArgPhe 490
   ::::: ||||| ::::: |||||
1392 AGCAGACAAAGCTGTTTCTTCTTCCAACTACAAGACGAGAAACGGTAC 1441
   ||||| ::::: |||||
491 ThrTyrPhe 493
   |||||
1442 CCGATCTTC 1450
   |||||
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC47986
seq_documentation_block:
ID AAC47986 standard; DNA; 1649 BP.
XX
AC AAC47986;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55845.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.

```

PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132663.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.

474 etlysrhrSerProThrValTyrProValaspasnleuProAlaargPhe 490
::: ::: ::::|:::|::: ||:::
1394 AGCAAGACAAGCTGGTTTCTTCCAACTACAGAAGCGCAGAAACGGTAC 1443
491 ThrTyrpHe 493
|||
1444 CCGATCTTC 1452

seq_name: /SIDSl/gcgcdata/geneseq/geneseqn/NA2000.DAT: AAC35582

seq_documentation_block:

ID AAC35582 standard; DNA; 1667 BP.

XX AAC35582;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10705.

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR	11-AUG-1999;	99US-0148319
PR	12-AUG-1999;	99US-0148341
PR	13-AUG-1999;	99US-0148565
PR	13-AUG-1999;	99US-0148684
PR	16-AUG-1999;	99US-0149368
PR	17-AUG-1999;	99US-0149175
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	20-AUG-1999;	99US-0149723
PR	20-AUG-1999;	99US-0149929
PR	23-AUG-1999;	99US-0149902
PR	23-AUG-1999;	99US-0149930
PR	25-AUG-1999;	99US-0150566
PR	26-AUG-1999;	99US-0150884
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153378
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154039
PR	22-SEP-1999;	99US-0154779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156458
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	08-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0159293
PR	13-OCT-1999;	99US-0159294
PR	13-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159638
PR	18-OCT-1999;	99US-0159584
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989
PR	25-OCT-1999;	99US-0161404
PR	25-OCT-1999;	99US-0161405
PR	25-OCT-1999;	99US-0161406
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161361
PR	28-OCT-1999;	99US-0161920
PR	28-OCT-1999;	99US-0161992
PR	28-OCT-1999;	99US-0161993
PR	29-OCT-1999;	99US-0162142

alignment_scores:		
Quality:	423.50	Length: 486
Ratio:	1.523	Gaps: 16
Percent Similarity:	57.202	Percent Identity: 27.984

```
alignment_block:
US-09-668-482-32 x AAC35582 ..
```

Align seg 1/1 to: AAC35582 from: 1 to: 1667

1	MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr	17
109	TTGGGCTCTTCTGTGATCATTTGTT..TGTTATGTACT.....	144
17	OLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS	34
145GCTCTTCTCCGATGGAAACCAGATGCGATATT	175
34	erSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly	50
176	CTAAGAAAGGT.....CTTCCTCTCGAACCATGGGC	207
51	PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgGlySpH	67
208	TGGCCAATATTGTGTGAACCACCTGAGTTCTTAACAAGACCATGATT	257
67	eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP	84
258	CATGAATAAACCAAGACTAAGATATGGGAGTTTCTCAAGTCTCACATT	307
84	heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle	100
308	TTGGCTGCCCAACAATAGTCTCAATGACGCGAGTTAAACAGATACATT	357
101	LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr	117
358	CTAATGAATGAATCGAAAGACTGTGTGCGGTTACCCGCAATCTATGCT	407
117	gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL	134
408	TGATATTCTAGGGACATGCAACATAGCTGCGGTTCATGGCCGCGACCAC	457
134	ysGlnArgLysLysValIleMetGlnAlaPheSerArgLysAlaLeuGln	150
458	GGCTAATGAGAGCGCTCGTTGCTTTCTTAATAAGCCCAACCATGAGAA	507
151	CysTyrValLeuValIleAlaGluGlnValSerSerCysLeuGlnGlnTr	167
508	GACCAATCTCTTG.....CCTAAGATTGATGATTTCATGAGAACTA	548
167	PLeuSerCysGlyGluArgGlyLeuLeu.....ValTyrProGluV	181
549	TCTT...TGCGGTTGGATGATCTTGAGACAGTTGATATCCAGAAAGA	595
181	AlLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu	197
596	CCAACATATGGCATTTTATCATCTGTTTACAAATAGCTGAGACTTTG	645
198	ProGlyProAlaGlyGlyGlyLysAspGluGlnGlnLeuValGluAlaPh	214
646	AAAAAACCA.....GAGGTTGAAGAAATATAGAACAGAGTTT	683
214	eGluGlnMetThrArgAsnLeuPheSerLeuProIleAspValProPhes	231
684	CAAGCTTGTGTGGGAACCTTA..TCGGTCCCGATCATATCCCGGAA	730
231	erGlyLeuTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIle	247
731	CGAATTACCGCAGTGGAGTCCAGCAAGCAAGAAC.....	762
248	GluGluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAs	264
763AACATCGATAGGTTATTGACAGAACTGATCAAGAAAGAAAGA	806
264	peGlyGly....CysLysAspAlaLeuGlnLeuLeuIleGluHisSerT	279
807	GTCGTGAGAAACTTTCACAGACATGTTGGGTTACTTGATGAAGAAGAAAG	856
279	rpGluArgGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThr	295

857 ATAACCGATACTGTGTTAACCGATAAAGAGATAAGAGATCAAGTGTAAACG 906
296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLe 312
:::||||:::|||||:::|||||:::|||||:::|||||:::|||||
907 .. ATCTGTATTCGCGTATGAGACTGTCTCTACACCTCCATGATGGC 953
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluG 329
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
954 TCTTAAGTATCTCCATGATCATCCAAAAGCTCTTGAAGAAGCTCAGAAGAG 1003
329 luIleLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 345
||:::|||||:::|||||:::|||||:::|||||:::|||||
1004 AA.....CATTTGGCTATTAAAGGAGAGAAAAACGACCTGACGAAACGCTC 1047
346 AspMetGluThrLeuGlnLeuLysTyrIleGlyCysValIleLysGlu 362
:::|||||:::|||||:::|||||:::|||||:::|||||
1048 ACCTTCGACGATATTAATATCGATGAATTCACCTCGAGCTGTGATCTTTGA 1097
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuL 379
||||||||||:::|||||:::|||||:::|||||:::|||||
1098 GACATCAAGATTGGCAACGATTGTTAATGGTGTCTTAGGAAACTACTC 1147
379 ysrThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnVal... 394
:::|||||:::|||||:::|||||:::|||||:::|||||
1148 ACGACTTGAACCTCAACGGTTATTATCCCAAAAGTTGGAGAAATTTC 1197
395 IleTyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLys 411
:::|||||:::|||||:::|||||:::|||||:::|||||
1198 GTATACACAAGAGAGATTAACTATGATACATCT...CTTTATGAGATGCC 1244
411 sGluGluPheAsnProAspArgPheIleValProHisProGluAspAla 428
|||||:::|||||:::|||||:::|||||:::|||||
1245 AATGATCTTTAACCCATGAGATGATGGAAGAGAGCTTAGAA.....T 1288
428 erArgPheSerPheIleProPheGlyGlyGlyLeuArgSerCysValGly 444
||:|:::|||||:::|||||:::|||||:::|||||
1289 CAAAGAGCTATTCTTACTCTTGGAGGTGAGATTAGGCTTGGCCCTGGA 1338
445 LysGluPheAlaLysIleLeuLeuLysIlePheThrValGluLeuAla 461
|||||:::|||||:::|||||:::|||||:::|||||
1339 AAGGACTAGGAATCTCGGAAGTCTCAAGCTTCCTTCACTACTTGTTC 1388
461 gHisCysAspTyrGlnLeuLeuAsnGlyProProThrMetLysThrSerP 478
||:|:::|||||:::|||||:::|||||:::|||||
1389 AAAATATAGATGGGAA...GAGAATGAGAGAACAATTAATGCTCTTTC 1435
478 roThrVal 480
|||
1436 CAAGAGTT 1443

OM of: US-09-668-482-32 to: Issued_Patents_NA:* out_format : pfs
Date: Nov 6, 2001 3:39 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+-p2n.model -DEV=xlp
-Q=/cgn2_1/USPRO_spool/US09668482/runat_05112001_133353_10897/app_query.fasta_1.1680
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09668482@cgn1_1_104 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-668-482-32
Query length: 497
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 149.780000

score_list:

Sequence	Strd Orig	Zscore	Escore	len	Documentation	..	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-08-882-164D-31	2612.00	5657.71	0.0	1725	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-724-466B-5	2437.00	5277.52	2.6e-286	1494	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-08-882-164D-5	2437.00	5277.52	2.6e-286	1494	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-08-882-164D-38	1853.00	3988.67	1.6e-214	4164	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-724-466B-3	1734.00	3739.68	1.2e-200	1850	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-08-882-164D-3	1734.00	3739.68	1.2e-200	1850	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-08-882-164D-36	1251.00	2680.12	1.2e-141	2677	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-882-164D-11	604.00	1294.56	1.9e-64	351	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-08-882-164D-11	604.00	1294.56	1.9e-64	351	!	
/cgn2_6/ptodata/2/ina/5B_COMB.seq	US-08-622-166A-1	450.00	937.97	1.4e-44	1608	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-08-991-677-3	358.50	736.07	2.4e-33	1883	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-091-432-1	337.00	689.45	9.4e-31	1838	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-103-840A-2	329.00	568.01	5.5e-24	4403765	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-103-840A-1	329.00	567.99	5.5e-24	4411529	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-103-840A-2	317.00	541.81	1.6e-22	4403765	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-103-840A-1	317.00	541.79	1.6e-22	4411529	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-948-564-15	300.00	608.71	3.0e-26	1831	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-948-564-3	285.00	577.02	1.7e-24	1691	!	
/cgn2_6/ptodata/2/ina/5A_COMB.seq	US-08-313-075A-49	270.50	544.85	1.1e-22	1757	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-292-768-67	262.50	528.10	9.1e-22	1665	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-881-784-8	261.50	525.91	1.2e-21	1665	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-292-768-3	261.50	525.91	1.2e-21	1665	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-292-768-69	261.50	525.91	1.2e-21	1665	!	
/cgn2_6/ptodata/2/ina/5A_COMB.seq	US-08-313-075A-29	258.50	519.36	2.8e-21	1665	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-948-564-5	255.50	512.98	6.3e-21	1644	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-948-564-1	255.50	511.49	7.7e-21	1838	!	
/cgn2_6/ptodata/2/ina/5B_COMB.seq	US-09-091-432-3	255.50	497.71	4.5e-20	5156	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-08-882-164D-37	255.00	523.63	1.6e-21	683	!	
/cgn2_6/ptodata/2/ina/5A_COMB.seq	US-07-912-900-24	250.50	501.89	2.6e-20	1666	!	
/cgn2_6/ptodata/2/ina/5A_COMB.seq	US-08-285-309-24	250.50	501.89	2.6e-20	1666	!	
/cgn2_6/ptodata/2/ina/5B_COMB.seq	US-08-502-046-24	250.50	501.89	2.6e-20	1666	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-606-505B-63	249.00	495.06	6.3e-20	2174	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-616-990-63	249.00	495.06	6.3e-20	2174	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-292-768-63	246.50	492.41	8.9e-19	1762	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-09-033-055A-1	243.50	486.16	2.0e-19	1722	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-08-881-784-5	241.50	481.49	3.6e-19	1762	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-292-768-1	241.50	481.49	3.6e-19	1762	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-292-768-65	241.50	481.49	3.6e-19	1762	!	
/cgn2_6/ptodata/2/ina/6A_COMB.seq	US-09-033-055A-3	241.00	480.40	4.1e-19	1761	!	
/cgn2_6/ptodata/2/ina/6B_COMB.seq	US-09-172-339-5	241.00	480.40	4.1e-19	1762	!	

/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-17 + 238.50 475.39 7.9e-19 1704
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-194-981E-3 + 237.50 474.12 9.3e-19 1591
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-948-564-9 + 237.50 472.56 1.1e-18 1788
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-201-118-10 + 237.50 472.09 1.2e-18 1852
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-238-821B-10 + 237.50 472.09 1.2e-18 1852

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-31

seq_documentation_block:

; Sequence 31, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1725 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-882-164D-31

alignment_scores:

Quality: 2612.00 Length: 497
Ratio: 5.256 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-668-482-32 x US-08-882-164D-31 ..

Align seg 1/1 to: US-08-882-164D-31 from: 1 to: 1725

1 MetcGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
25 ATGGGGCTCCCGGGCTGCTGGCCAGTGGCGCTGCACCTTCGTGCTGCC 74
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrcysValS 34
|||||
75 GCTGCTGCTCTTCCTGGCGGCGCTCAAGCTCTGGGACCTGTACTGTGTA 124
34 erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
125 GCAGCCCGCATCGCAGCTGCGCCCTCCCTTGCCCCCGGTACCATGGGC 174

51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
175 TTCCCATTTCTTGGGGAACATTCAGATGGTCTTCAGCGGAGGAAGTT 224
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
225 TCTGCAGATGAGCGCAGGAATACGGCTTCATCTACAAGACGCATCTGT 274
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
275 TTGGCGCGCCACGGGTGCGGGTGATGGCGCGGATATATGTGCGGCGCATC 324
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
325 TTGCTGGAGAGACACCGGTGTGTGGTGCACTGGCCCGCGTGGTGG 374
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
375 CACCATCTGGGGCGCTGGCTGCCCTCCCAACCTGCACGATTCCTCGCACA 424
134 ySGlnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
|||||
425 AGCAGCGAAGAAGGTGATATATGACAGGCTTCACCGCGAGGCACATCCAG 474
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGluInTr 167
|||||
475 TGCTACGTGCTCGTGATCGCTGAGGAGATCAGCAGTGTGTGAGACAGTG 524
167 pLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
525 GCTAAGCTGGCGGCGAGCGCGCCCTCTGTGTTCTACCCCGAGGTGAAGCGCC 574
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
|||||
575 TCATGTTCCGCATCGCCATGCGCATCTGCTGGGCTGCGAGCGGGTCCA 624
201 AlAGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGluMe 217
|||||
625 GCGGCGCGCGGGGAGGAGCAGACAGCAACACTGCTGGAGGCTTTCGAGGAGAT 674
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
675 GACCCGCAATCTCTTCTCTCTCCATTCGACGTGCCCTTAGCGGCTGT 724
234 yTArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
|||||
725 ACCGGGCGGTGAAGCGCGGAACCTTATACACGCGGCATCGAGAGAAAC 774
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
|||||
775 ATTGCGGCCAAGATCCGCGCGCTTCAGGCTACAGAGCCGGATGGGGGTTG 824
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
|||||
825 CAAGGACGCGCTGCAGCTCCTGATTGAGCACTCGTGGGAGAGGGAGAGA 874
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheGly 300
|||||
875 GGCTGGATATGCAGGCACATAAACATCGTCAACAGAGCTCTCTTTGGT 924
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
|||||
925 GGTCATGAAACTACAGCCAGTGTGCGACATCATGATCACTTAACCTAGG 974
317 yLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLysG 334
|||||
975 ACTTACCCACATGTCTCTCCAGAAAGTTGAGAGAAGATTAAGACCAAGG 1024
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
|||||
1025 GCTTACTTTGCAAGAGCAATCAAGACAACAAGTTAGACATGAAAACTTTG 1074

351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1075 GAACAGCTTAATACATTTGGGTGTGTGCATTAAGAGAACCCCTGCATTGAA 1124
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1125 TCCTCCGGTTCAGAGGGGTTTCGGGTGCTGTAAGACTTTTGAGCTGA 1174
384 sngLYTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
|||||
1175 ATGATACCAGATCCCCAAGGGCTGGAATGTTATTACAGTATCTGTGAC 1224
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1225 ACCCAGATGTGGCAGATATCTTCACCTAACAGAGGAATTAATCCCGA 1274
417 pArgPheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||
1275 CCGCTTATAGTGCTCATCCAGAGGATGTTCCCGGTTCAGCTTCATTC 1324
434 ropheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1325 CATTTGAGAGAGCCTTCGAGAGCTGTGTAGGCAAAAGTTGCCAAAATT 1374
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
|||||
1375 CTTCCTTAAGATATTACAGTGAGCTGGCTAGGCACGTGATTTGGCAGCT 1424
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1425 TCTAATGAGACTCTCAATGAAGACAGAGCCCACTGTGTACCCCTGTGG 1474
484 spAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||
1475 ACAATCTCCCTGCAAGATTCACTACTTCCAGGAGAGATATC 1515

seq_name: /cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-724-466B-5

seq_documentation_block:

: Sequence 5, Application US/08724466B

: Patent No. 6063606

: GENERAL INFORMATION:

: APPLICANT: Petkovich, P. Martin, White, Jay A.,

: APPLICANT: Beckett, Barbara R., Jones, Glenville

: TITLE OF INVENTION: Retinoid Metabolizing Protein

: NUMBER OF SEQUENCES: 30

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Blake, Cassels & Graydon

: STREET: Box 25, Commerce Court West

: CITY: Toronto

: ZIP: M5L 1A9

: COUNTRY: Canada

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage

: COMPUTER: COMPAQ, IBM PC compatible

: OPERATING SYSTEM: MS-DOS 5.1

: SOFTWARE: WORD PERFECT

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/724,466B

: FILING DATE: October 1, 1996

: PRIORITY APPLICATION DATA:

: APPLICATION NUMBER: 08/667,546

: FILING DATE: June 21, 1996

: ATTORNEY/AGENT INFORMATION:

: NAME: Hunt, John C.

: REGISTRATION NUMBER: 36,424

: REFERENCE/DOCKET NUMBER: 50767/00004

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (416) 863-4344

: TELEFAX: (416) 863-2653

: INFORMATION FOR SEQ ID NO: 5:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1494 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-724-466B-5

alignment_scores:
Quality: 2437.00 Length: 497
Ratio: 5.035 Gaps: 0
Percent Similarity: 97.384 Percent Identity: 93.360

alignment_block:
US-09-668-482-32 x US-08-724-466B-5 ..

Align seg 1/1 to: US-08-724-466B-5 from: 1 to: 1494

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
1 ATGGGGCTCCCGCGCTGCTGCCAGTGCGGCTGCACTTCGTGCTGCC 50
17 OleuLeuLeuPheLeuAlaLeuLysLeuTrpAspLeuTyrCysValS 34
51 GCTGCTGCTCTTCTCTGGCTGGCATCAAGCTCTGGGACCTGTACTGCGTGA 100
34 eSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
101 GCGCGCGCAGCCGAGTTGTGCTCCCTCCATTGCCCCCGGGACTATGGGC 150
51 PheProPhePheGlyGlyThrLeuGlnMetValLeuGlnArgArgLysPh 67
151 TTCCCTCTCTTTGGGAAACCTTGCAGATGGTACTGACGCGAGGAAGTT 200
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
201 CCTGCAGATGAAGCGCAGAAATACGGCTTCATCTACAAGACGCAATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
251 TCGGGCGGCCACCGCTACGGGTGATGGCGCGGACAATGTGGCGGCATC 300
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
301 TTGCTCGGAGACGACCGGCTGGTGCCTCCACTGCGCAGCGTCGTCGG 350
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
351 CACCATTTCTGGGATCTGGCTGCCTCTTAACCTGCACGACTCCTCGCACA 400
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
401 AGCAGCGCAAGAGGTGATTATGGGGCTTCAGCCGCGAGGCACCTCGAA 450
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGlnIleTr 167
451 TGCTAGTGGCGGTGATCACCAGAGAAAGTGGCAGACCTGGAGCAGTG 500
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
501 GCTGAGCTGCGCGGAGCGCGGCTCCTGTTCTACCCCGAGGTGAAGCGCC 550
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
551 TCATGTTCGGAATCGCATGCGCATCTTACTGGGCTGCGAACCCTCACTG 600
201 AlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGluMe 217
601 GCGGGCGAGCGGGACTCCGAGCAGCAGAGCTTGTGAGAGGCTTCGAGGAAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
651 GACCCGCAATCTCTTCGCTGCCCATCGACGTGCCCTTCAGCGGGGCTGT 700
234 yrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250

|||||:|||||
701 ACCGGGGCATGAAGCGCGGAACCTCATTCACGCGCGCATCGACAGAAC 750
251 IleArgAlaIleLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
751 ATTCGGCGCCAGATCTGCGGGCTGGCGCATCCGAGCGCGGCCAGGGCTG 800
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGluA 284
801 CAAGACGCGCTGCAGCTGTGTATCGACACTCGTGGAGAGGGGAGAGAC 850
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
851 GGCTGACATGACAGGCACTAAAGCAATCTCAACCGAATCCTCTTTGGA 900
301 GlyHisGlyThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG1 317
901 GGACACGAACACCAAGCGCCAGTGCAGCACATCTCTGATCATTACTTGGG 950
317 yLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLysG 334
951 GCTTACCCACATGTCTTCCAGAAAGTGCAGAAAGAGCTGAAGAGTAAAG 1000
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
1001 GTTACTTTGCAAGACCAATCAAGACACAGATTGGACATGGAATTTTG 1050
351 GluGlnLeuLysTyrIleGlyCysValIleLysGlyThrLeuArgLeuAs 367
1051 GAACAACTTAATACATCGGGTGTGTTATTAAAGAGACCCTTCGACTGAA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
1101 TCCCCAGTTCCAGAGGGGTTTGGGTTGCTCTGAAGACTTTTGAATTAA 1150
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
1151 ATGGATACCAAGATTCGCCAAGGGCTGGAATGTATCTACAGTATCTGTGAT 1200
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
1201 ACTCATGATGTGGCAGAGATCTTCACCAACAGAGAAATTTAATCCTGA 1250
417 parGpHeIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
1251 CCATTTCACTGCTCCTCACCACAGAGATGCATCCAGTTCAAGTTCAATTC 1300
434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1301 CATTTGAGGAGGCGCTTAGAGCTGTGTAGGCAAGAAATTTGCCAAAATTT 1350
451 leuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
1351 CTTCCTCAAAATATTATTACAGTGAAGCTGGCCAGGCATTTGACTGGCAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
1401 TCTAATGGAACCTCTACAATGAAGAAACAGTCCACCGCTGTATCCTGTGG 1450
484 spAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
1451 ACAATCTCCCTGCAAGATTCAACCAATTTCCATGGGGAATTC 1491
seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-882-164D-5
seq_documentation_block:
: Sequence 5, Application US/08882164D
: Patent No. 6306624
: GENERAL INFORMATION:
: APPLICANT: Petkovich, P. Martin, White, Jay A.,
: APPLICANT: Beckett, Barbara R., Jones, Glenville
: TITLE OF INVENTION: Retinoid Metabolizing Protein
: NUMBER OF SEQUENCES: 43

```

CORRESPONDENCE ADDRESS:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667, 546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1494 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-5

alignment_scores:
Quality: 2437.00          Length: 497
Ratio: 5.035              Gaps: 0
Percent Similarity: 97.384 Percent Identity: 93.360

alignment_block:
US-09-668-482-32 x US-08-882-164D-5 ..

Align seg 1/1 to: US-08-882-164D-5 from: 1 to: 1494

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
1 ATGGGGCTCCCGCGCTGCTGGCCAGTGCCTCTGCACCTTCGTGCTGCC 50
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
|||||
51 GCTGCTGCTCTTCCTGCTGCGATCAAGCTCTGGGACTGTACTGCGTGA 100
34 erSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
101 GCGGCGCGGACCGCAAGTTGTGCCCTCCCATTTGCCCCCGGGACTATGGGC 150
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
151 TTCCTCTTCTTTGGGAAACCTTGCAGATGTTACTGCAGCGAGAAGTT 200
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
201 CCTGCAGATGAAGCGCAGGAATAACGGCTTCATCTACAGACGCATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
251 TCGGGCGGCCACCGCTACGGGTGATGGCGCGGACAATGTGCGGCGCATC 300
101 leuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
301 TTGCTCGAGACGACCGGCTGGTGTGCTCCACTGGCCAGCGTGGTGG 350
```

```

117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
351 CACCATTTCTGGGATCTGGCTGCCTCTCTAACCCTGCACGACTCCTCCACA 400
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
|||||
401 AGCAGCGCAAGAGGTGATATTATGCGGGCTTCAGCCGCGAGGCACTCGAA 450
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGlnGlnTr 167
|||||
451 TGCTACGTGCGGCTGATCACCGAGGAAGTGGCAGCAGCTGGAGCAGTG 500
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
501 GCTGAGCTGCGCGGAGCGCGGCTCCTGTCTACCCCGAGGTGAAGCGCC 550
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
|||||
551 TCATGTTCCGAATCGCCATGCGCATCTACTGGGCTGCGAACCACAAC TG 600
201 AlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGluMe 217
|||||
601 GCGGGCGAGCGGGAAGCTCCGAGCAGCAGCTTGTGAGGCTTCAGAGAAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
651 GACCGCAATCTTCTCTCGCTGCCATCGACGTGCCCTTCAGCGGCTGT 700
234 yTrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
|||||
701 ACCGGGCGCATGAAGCGCGGACCTCATTCACGCGCGCATCGACAGAAAC 750
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
|||||
751 ATTGCGGCCAAGATCTGCGGGCTGCGGCAATCCGAGGCGGCGGCGGCTG 800
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTyrGluArgGlyGluA 284
|||||
801 CAAGACGCGCTGCAGCTGTTGATCGACACTCGTGGAGAGGGAGAGAGC 850
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
851 GGCTGGACATGCAGGCACCTAAAGCAATCTTCAACCGAATCCTCTTTGGA 900
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
|||||
901 GGACACGAAACACAGCGGCTGACGACACCATCTCTGATCACTTACCTGGG 950
317 yLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLysG 334
|||||
951 GCTTACCCACATGTTCTCAGAAAGTGCAGAGAGGCTGAAGAGTAAGG 1000
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
|||||
1001 GTTACTTTTGAAGAGCAATCAAGACAACAAGTTGGACATGGAATTTTG 1050
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1051 GAACAACCTTAATACATCGGGTGTGTATTAAAGAGACCCCTTCGACTGAA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1101 TCCCCCAGTTCAGAGAGGCTTCGGGTTGCTCTGAAGACTTTTGAATTAA 1150
384 sngLyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
|||||
1151 ATGCATACCAAGATTCACAGGCGCTGGAATGTTATCTACAGTATCTGTGAT 1200
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1201 ACTCATGATGTGGAGAGATCTTCAACCAAGAGAAATTTAATCTCTGA 1250
```



```
417 pargpheileValProHisProGluAspAlaSerArgpHeserPheilep 434
||||| :|||
1251 CGATTCACTGCTCTCCACCCAGAGATGCATCCAGTTCAGCTTCATTC 1300
434 ropheglglyglyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1301 CATTTGAGAGAGCCCTAGAGCTGTGTAGGCAAGAATTTGCCAAAATT 1350
451 LeuleulysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnIle 467
|||||
1351 CTTCCTCAAAATATTATTACAGTGGAGCTGGCCAGGCATTTGTGACTGGCAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1401 TCTAAATGGACCTCTTACAAATGAAAACCAAGTCCACCGCTGTATCTGTGG 1450
484 spAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||
1451 ACAATCTCCCTGCAAGATTCACCAATTTCCATGGGGAATC 1491
```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-882-164D-38

seq_documentation_block:

```
; Sequence 38, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4164 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-882-164D-38
```

alignment_scores:

```
Quality: 1853.00      Length: 1046
Ratio: 4.240          Gaps: 10
Percent Similarity: 41.778      Percent Identity: 41.396
```

alignment_block:

US-09-668-482-32 x US-08-882-164D-38

Align seg 1/1 to: US-08-882-164D-38 from: 1 to: 4164

```
40 CysAlaLeuProLeuProProGlyThrMetGlyPheProPhePheGlyI 56
|||||
73 TGGGCTCTCTCTCTCCACCTCCCTCT..... 99
56 uThrLeuGlnMetValLeuGlnArgArgLysPheLeuGlnMetLysArg 73
100 .....CAGCGAGGAGTTTCTGCAGATGAAGCGCA 130
73 rGlySTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVal 89
|||||
131 GGAATACGGCTTCATCTACAGACGCAATCTTTGGCGGCCACCGTG 180
90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyGluHisAr 106
|||||
181 CGGGTATGGGCGGATATATGTGGCGGCATCTTGCTGGAGAGCACCG 230
106 gLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAlaG 123
|||||
231 GTTGTGTGCGTGCACTGGCCCGCTCGTGCGCACCATCTTGGCGCGCTG 280
123 lYcysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys..Lys.. 138
|||||
281 GCTGCTCTCCAACTGCACGATTCCTCGCAAGACGAGAAAGAGGTG 330
138 ..... 138
331 AGGTGAGCTGGCAACTCTTGCTGGCTGGCAGGAGACCTCATCTATGGCT 380
138 ..... 138
381 TGGTTACGGCAAAATAGAAATGCGGGCGAGGCTAGTCTATGTGTGGG 430
138 ..... 138
431 GACCAGACCCTCTCTATCTGAGATCCACTTACCTTTCTGCTAGCACG 480
138 ..... 138
481 TGGTTAGTCTGCGGGGAGCTGAATTTCTGAAGGGTACTCGGAAGG 530
138 ..... 138
531 CGAAGGGGGGGGCTGAGGGAAGTAGAGATTGTAACACTCTCTGCTC 580
139 .....ValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
581 CTGGGGGTGCTCAGTGATTTATGACAGGCTTCAGCCGCGAGGCACTCCA 630
150 nCysTyrValLeuValIleAlaGluGluValSerSerCysLeuGluGlnT 167
|||||
631 GTGCTACGTGCCCGTATGCTGAGGAAGTCAGCACTTGTCTGAGCAGT 680
167 rPLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183
|||||
681 GGCTAAGCTGCGGCGAGCGCGGCTCTGCTACCCCGAGGTGAAGCCG 730
184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPr 200
|||||
731 CTCATGTTCCGATCGCCATGCGCATCTGCTGGGCTGCGAGCCGGGTCC 780
200 oAlaGlyGlyGlyAspGluGlnGlnLeuValGluAlaPheGluGluM 217
|||||
781 AGCGGGCGGGGAGAGCAGACGACAGCTGTGAGAGGCTTTGAGGAGA 830
217 eThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
|||||
831 TGACCCGCAATCTCTCTCTCTCCATGACGTGCCCTTTAGCGGCGCTG 880
234 TyrArg..... 235
```

```
881 TACCGGGTAAGGGCGGTTTGGGAGTCGAGTAGGGGAAACGCAAGCTCGG 930
236 .....Gly.ValLysAlaarg 240
931 GCATCCGCTACCGCCACGCTCTCTCCGCGCTCAGGGCGGTGAAGCGCGG 980
241 AsnLeuIleHsIlaargIleGluIaasnIleargAlaLysIleargar 257
|||||
981 AACCTTATACAGCGCGCATCGAGAGAACATTCGCGCAAGATCCGCCG 1030
257 glenGlnAlaThrGluProaspGlyIcysIysaspAlaLeuGlnLeu 274
|||||
1031 GCTTCAGGCTACAGACCGGATGGGGGTGCAAGACCGCTGCAGCTCC 1080
274 euIleGluHisSerTrpGluArgGlyGluArgLeuAspMetGln.... 288
|||||
1081 TGATTGAGCACTCGTGGAGAGAGGGGAGAGAGGCTGGATATGCAGGTGAGA 1130
288 ..... 288
1131 AGCAATTTCAAAAGGTGCCAAGGGCCGGGAGTGCCCTTGACTTTCAGA 1180
288 ..... 288
1181 CACACTTTCTGGGGTCTCCAAAGCCCTGTCAAGGCCCCAGCTACTTCCAA 1230
288 ..... 288
1231 GTGGGGCGGATGCTAGGTCTAGAGCTTTTCAACCTGTGGTGTGACCC 1280
288 ..... 288
1281 CTTACGAGGCAACAACCCCTTTCAGAAGGTCGCCCTAAGAGCATCTGC 1330
288 ..... 288
1331 ATATCCGATATTTACATCAAGAAACATACAGTAGCAAAATTACCGTTAT 1380
288 ..... 288
1381 GAAGTAGCAACAAGATAATTATTATCGTTGGGGGTCAACCAACACGAGG 1430
288 ..... 288
1431 AACCGTATTAAAGGGGTGCATTTGCTTAGAGAGCTGTGGAAGGGGTGGC 1480
288 ..... 288
1481 TGAGCAATGGGGAAGATCCCAAAGTTCAAAGGGCAAGCTCATCTACAAA 1530
288 ..... 288
1531 GGTTAAAGCGGAAGACGAGATTAAAGGAGTTTGGCGTTTGTGTGTGG 1580
288 ..... 288
1581 TCTTTGACTTTCTATGAACAACCGATTTTACCCCTGAAGTCTTCCGTG 1630
288 ..... 288
1631 CAATATTCTCAGGTCAGGTCCTTTGTACAGTGCTATAAACTGCACCTCAGA 1680
289 .....AlaLeuLysGlnSerSerThr 295
1681 TCTGTATTAACCTTCGGTTTATTCCTTAGGCACATAAAACAATCGTCAACA 1730
296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLe 312
|||||
1731 GAGCTCCTCTTTGGTGTGATGAACACTACAGCCAGTGTGCGACGTCACT 1780
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluG 329
|||||
1781 GATCAGCTTAACCTAGGACTCTACCCACATGTCTCCAGAAAGTTCGAGAAG 1830
```

```
329 IuIleLysSer..... 332
|||||
1831 AGATAAAGAGCAAGGTAGATGATTTCTAGAGGTTCCCATTTGCCTAGGA 1880
332 ..... 332
1881 CATTCCTCTATTAAACCAACCACCACCACCCCACTGTATATAAGTTGCTC 1930
332 ..... 332
1931 GATACACCAGTACTATGACAGTGAAGATCTGAGAGCTAGGTGGACTGT 1980
332 ..... 332
1981 GGGGAGAGACTCCACCTCGTGAATTTAAAAAGGAGTTGTTGTACTGG 2030
332 ..... 332
2031 GCTCTCTCTTGGGCAAAATTGACCCCTCTCCCTCTCCTCCTCCTCCT 2080
332 ..... 332
2081 CCTCTCTCCTCCACCAACCACCACCACCATCACCACCTTTATAGAGCAAG 2130
332 ..... 332
2131 TTCTCCTTCCCTGACCAAGAACATGAAATGATGATTAGAGCAATAGC 2180
332 ..... 332
2181 TGATCAGGTCGCAGTGTGTGAGGGCTCAGGGTATGACCCTTATATA 2230
332 ..... 332
2231 CCTGATAAGCAACATTGCTCGATAATGGGTTAGGCTGAGGAAGTGTG 2280
332 ..... 332
2281 AAAGGAAGGCCATCAGGCCATCAGCTCTTCCCTTTATCCTCCCATC 2330
332 ..... 332
2331 CAGACGCCCTTACGTTTAGTTAACAGGTGAGTCTGCTGGGCTGACTTTT 2380
333 .....Lys. 333
2381 TTTTGGAGTGCCAGGGATCCATCAGTCACCTTTTATCTGTTCATA 2430
334 .GlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrL 350
|||||
2431 GGGCTTACTTTGCAAGACAAATCAAGACAACAAGTTAGACATGAAACTT 2480
350 euGluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeu 366
|||
2481 TGGCACAGCTTAATATACACTGGGTGTGTCAFTTAAGAGACCCGCAATTG 2530
367 AsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu.. 382
|||||
2531 AATCCTCCGGTTCAGAGAGGGTTTCGGGTGCTCTGAAGACTTTTGAGCT 2580
382 ..... 382
2581 GAATGTGAGTGACCTCCTGTCCCCCACCACCCCAAGCCCTGTCCACGTC 2630
382 ..... 382
2631 CTCTGTATGTGTGAGCATCAGCTGCCAGAGCAGTGGCTCACTGCC 2680
382 ..... 382
2681 TTGACAGTGTCCTGCCTCTATGTGTAAGTGGGAACCAATTTGCTCTCCT 2730
```

```
382 ..... 382
2731 CTTAATGCCATCATGCTAGTAATGACTTTTGTGTGTCAGAGCTCAGGG 2780
383 .....LeuAsn..GlyTyrG 387
2781 CCGGGAATGTCAATCTTAGAATTTTTTTTTTTTAAACAGGATACC 2830
387 InIleProLysGlyTrpAsnValIleTyrSerIleCysAspThrHisAsp 403
2831 AGATCCCAAGGGCTGGAATGTTATTACAGTATCTGTGACACCCAGCAT 2880
404 ValAlaAspIlePheThrAsnLysGluGluPheAsnProAspArgPheI 420
2881 GTGGCAGATATCTTCACTAACAGAGGAATTTAATCCCGACCGCTTTAT 2930
420 eValProHisProGluAspAlaSerArgPheSerPheIleProPheGlyG 437
2931 AGTGCCCTCATCCAGAGATGCTCCCGGTTCAAGCTTCATTCATTTGGAG 2980
437 IyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLeuLys 453
2981 GAGGCTTCGAGAGCTGTGTAGGCAAGAGATTGGCAAAATCTTCTTAAG 3030
454 IlePheThrValGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnG 470
3031 ATATTACAGTGGAGCTGGCTAGGCACTGTGATTGGCAGCTTCTAAATGG 3080
470 yProProThiMetLysThrSerProThrValTyrProValAspAsnLeuP 487
3081 ACCTCCTACAAATGAAGACAGACCCCACTGTGTACCCCTGTGACAACTCC 3130
487 roAlaArgPheThrTyrPheGlnGlyAspIle 497
3131 CTGCAAGATTACCCCACTTCCAGGAGATATC 3162
seq_name: /cgn2_6/prodata/2/ina/6A_COMB.seq:US-08-724-466B-3
seq_documentation_block:
; Sequence 3, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1850 base pairs
```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-724-466B-3

alignment_scores:
    Quality: 1734.00      Length: 496
    Ratio: 4.070          Gaps: 4
    Percent Similarity: 85.887      Percent Identity: 67.742

alignment_block:
US-09-668-482-32 x US-08-724-466B-3 ..
Align seg 1/1 to: US-08-724-466B-3 from: 1 to: 1850

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
   ||||| ..... |||||
88 ATGGGGCTGTACACCTTATGTTGTCACCTTTCTCTGCACCATCGTGTACC 137
17 oleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
   |:::||||| ..... |||||
138 CGTTTACTCTTCTCCGCCCGGTGAAGTTGTGGAGATGTTAATGATCC 187
34 erSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
   ||| ::||| |||||
188 GACGAGTCGATCCGAACGTCAAGAAGTCCCTACCGCAGGTACCATGGGC 237
51 pheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
   ::||| ..... |||||
238 TTGCCGTTCAATTGGAGAAACGCTCCAGCTGATCTCCAGAGAAGAAGATT 287
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
   ||||| ..... |||||
288 TCTGCCGATGAACGGCAGAAATACGGGTGCATCTACAGAAGCACCTCT 337
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
   ||||| ..... |||||
338 TCGGGAACCCGACTGTCAAGGGTGATGGAGCTGATTAATGTGAGGCAGATT 387
101 leuLeuGlyLunHisArgLeuValSerValHisTrpProAlaSerValAr 117
   ||||| ..... |||||
388 CTGCTGGGCGAACAACAGCTGGTGTCTGTTCAGTGGCCAGCATCAGTGAG 437
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
   ||||| ..... |||||
438 AACCATCCTGGGCTGTGACACCCCTCTCCAATGTCCATGAGATTCAACACA 487
134 ysgLInArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
   ||::: ..... |||||
488 AAAACAAGAAAAAGGCCATATGAGGGCGTCTCTCGAGATGCTCTGGAG 537
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGlnInTr 167
   ||::: ||||| ::|||
538 CACTACATTCCTCCGTGATCCAGCAGAGGTTGAAGAGCGGCCCATACAGAATG 587
167 pleu.....SerCysGlyGluArgGlyLeuLeuValTyrProGluV 181
   |||| |::| ..... |||||
588 GCTGCAAAAAGACTCCTGC.....GTGCTGCTTATCCAGAAA 625
181 alLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
   ::||:: ..... |||||
626 TGAAGAAACTCATGTTTCGATAGCTATGAGAAATCCTGCTGGTTTGA 675
198 ProGlyProAlaGlyGlyLysGluAspGluGlnGlnLeuValGluAlaPh 214
   ||| |::| ..... |||||
676 CCAGAGCAATA.....AAGAGGAGCAGAGCAAGAACTGTGGAAGCTTT 719
214 eGluGlnMetThrArgAsnLeuPheSerLeuProIleAspValProPheS 231
   ||||| ..... |||||
720 TGAAGAAATGATCAAAAACCTGTCTCTCCATGCAATGAGAGTTCCTTTCA 769
231 erGlyLeuTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIle 247
```

```

770 GTGTCTGTACAGGGGTTTGAGGGCGACGCCAATTCATTCTCACTCCAAATT      819
248 GluGluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAs       264
   |||||              |||||::: |||||::: |||||::: |||||:::
820 GAGGAATAACATCAGGAAGAAAATTCAGAT..GACGACAATGAAGAACA        866
264 pGlyGlyCysLysAspaLaLeuGlnLeuLeuIleGluHisSerTrpGluA     281
   :             |||||              |||||::: |||||::: |||||:::
867 ACAGAAATACAAAGACGCCCTTCAGCTGTGCATCGAAGAACAGACAGAGAA    916
281 rGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeu     297
   |||           ::::::::::::::::::::|||::: |||||::: |||||:::
917 GTGACGAACCTTTAGTTTGACGGCATGAAGAAGACAGCTACAGACTT       966
298 LeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleTh     314
   |||||          ::::::::::::::::::::|||::: |||||::: |||||:::
967 CTATTGTGAGGTCATGAACCACCGCCAGCACTGC AACCTCACTTGTCA      1016
314 rTyrlLeuGlyLeuTyrrProHisValLeuGlnLysValArgGluIleL      331
   ::::::::::|||           ::|||::: |||||::: |||||::: |||||:::
1017 GTTTCGTGGGTCTGACACACAGAGTGTGCAGAAGGTCAGAGAGAGGTTT    1066
331 ySserLySGlyLeuLeuCysLysSeraSnGlnAspaNlysLeuAspMet     347
   :::::|||           :::         :::         :::         |||||:::
1067 AGCAGAAGGTTGAATGGCGCATGTATACACCTGGAAGGGCTTGAGTATG    1116
348 GluThrLeuGluGlnLeuLysTyrrIleGlyCysValIleLysGluThrL     364
   |||  |||:::|||||          |||||          |||||          |||||
1117 GAGCTGTTGACACGAGCTGAAGTAACACTGGATGTGTGATTAAGACACTCT    1166
364 uArgLeuAsnProProvalProGlyGlyPheArgValAlaLeuLysThrp     381
   |||||:::|||||          |||||          |||||          |||||
1167 TAGAATCAACCCCTCCTGTTCGCCGAGGATTCAGAGTGCACCTCAAACCT    1216
381 heGluLeuAsnGlyTyrrGlnIleProLysGlyTrpAsnValIleTyrrSer   397
   |||||          |||||          |||||          |||||          |||||
1217 TTGAATTGAATGTTACCAAAATTCCTAAAGATGGAACGTCAATTACAGC    1266
398 IleCysAspThrHisAspValAlaAspilePheThrAsnLysGluGluPh     414
   |||||          |||||          |||||          |||||          |||||
1267 ATCTGTGACACGCACGATGTGCCCGCACGTCCTTCCAACAAGAAGAGATT    1316
414 eAsnProAsparGpheIleValProHisProGluAspaLaSerArgPheS     431
   :::::|||:::|||||:::         |||||::: |||||::: |||||:::
1317 CCAGCCGGAGAGATTCA TGAGCAAAAGCTCTGGAGAGCGGCTCCAGGTTT    1366
431 erPheIleProPheGlyGlyLeuArgSerCysValGlyLysGluPhe       447
   :::::|||          |||||          |||||          |||||          |||||
1367 ACTACATCCCCTTCGGAGGAGATCCAGAGATGTGTGGCAAAAGAGTTCC    1416
448 AlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAs     464
   |||||:::|||||          |||||          |||||          |||||          |||||
1417 GCCAAAGTGTACTCAAGATCTTTTAGTTGAGTTAACGCACGACTTGCAA    1466
464 pTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrValT     481
   :|||  |||  |||||          |||||          |||||          |||||          |||||
1467 TTGGATTCTCTCAACGAGACCCCCGACAATGAAGAAGCCCGCACTATT    1516
481 yrProValAspaSnLeuProAlaArgPheThrTyrrPhe 493
   |||||          |||||          |||||          |||||          |||||
1517 ACCCAGTGGACAATCTCCCTACCAAAATTCAC TACTTAT 1554

seq_name: /cgn2_6/pdata/2/ina/6B_COMB.seq:US-08-882-164D-3
seq_documentation_block:
; Sequence 3, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
```

```

CORRESPONDENCE ADDRESS:
ADDRESSSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 Inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEO ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1850 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

```
alignment_scores:
    Quality: 1734.00
    Ratio: 4.070
    Percent Similarity: 85.887
    Length: 496
    Gaps: 4
    Percent Identity: 67.742
```

alignment_block:

US-09-668-482-32 x US-08-882-164D-3

Align seg 1/1 to: US-08-882-164D-3 from: 1 to: 1850

1	MetGlyLeuP	ProAlaLeuLeuAlaSerAlaLeu	CysThrPheValLeu	Pr	17
		:::		:::	
88	ATGGGCGCTGTACACCCCTTATGGTCACTTCTCTGCA	CCATCGTCTACC			137
17	oleuLeuLeuPheLeuAlaAlaLeu	LysLeuTrpAspLeu	TyrCysVal	34	
				:::	
138	CGTTTACTCTTCTCTCGCCGGGTGAAGTTGTGGAGATGTTAATGATCC				187
34	erSerArgAspArgSerCysAlaLeu	ProLeuProProGlyThrMetGly		50	
		:::			
188	GACGATTCGATCCGAACTGCAGAGTCCCTCTACCGCAGGTACCATGGGC				237
51	PheProPhePheGlyGluThrLeu	GlnMetValLeuGlnArgGly	Sp	67	
	:::	:::	:::		
238	TTCGCCGTTCATTGGAGAAACGCTCCAGCTGATCCTCCAGAGAGGAAGTT				287
67	eLeuGlnMetLysArgArgGly	TyrGlyPheIleTyrLysThrHisLeu	P	84	
288	TCTGGGCATGAACCGGCAGAAATACGGGTGCATCTACAGACGCACCTCT				337
84	heGlyArgProThrValArgVal	MetGlyAlaAspAsnValArg	GlyIle	100	
338	TGCGGAACCCGCACTGTACAGGTGATGGAGCTGTATATGTGAGGCGAGATT				387
101	LeuLeuGlyGluHisArgLeuValSerValHis	TrpProAlaSerVal	Arg	117	
388	CTGCTGGCGGCAACACCAAGCTGTGTCTGTTCAGTGGCCAGCATCAGTGAG				437


```
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisaspSerSerHisL 134
|||||
438 AACCATCCCTGGGCTCTGACACCCCTCTCCATGTCCATGAGAGTTCAACACA 487
|||||
134 ySgInArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
|||||
488 AAAACAGAAAAAGGCCATATGAGGGCGTCTCTCGAGATGCTCTGGAG 537
|||||
151 CysTyrValLeuValIleAlaGluValSerSerCysLeuGluInTr 167
|||||
538 CACTACATTCCTCCGTGATCCAGCAGGAGGTGAAGAGCCCATACAGGAATG 587
|||||
167 pLeu.....SerCysGlyGluArgGlyLeuLeuValTyrProGluV 181
|||||
588 GCTGCAAAAAGACTCCTGC.....GTGCTGTTTATCCAGAAA 625
|||||
181 allYsArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
|||||
626 TGAAGAAACTCATGTTTCGATAGCTATGAGAATCCTGCTGGTTTGGAA 675
|||||
198 ProGlyProAlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPh 214
|||||
676 CCAGAGCAATA.....AAGACGAGACGACGAAGAACTGGTGAAGCTTT 719
|||||
214 eGluGluMetThrArgAsnLeuPheSerLeuProIleAspValProPheS 231
|||||
720 TGAGGAAATGATCAAAACTGTCTCTTCCATTCGACGTCCTTTCA 769
|||||
231 eRgIleuTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIle 247
|||||
770 GTGCTCTGTACAGGGGTTTGAGGGCAGCAATTTCATTCACCTCAAAAT 819
|||||
248 GluGluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAs 264
|||||
820 GAGGAAACATCAGAGAAATTCAGAT..GACGACAATGAAACGA 866
|||||
264 pGlyGlyCysLysAspAlaLeuGlnLeuIleGluHisSerTrpGluA 281
|||||
867 ACAGAAATACAAAGACGCCCTTCAGCTGTGATCGAGAACGACAGAGAA 916
|||||
281 rGgIyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeu 297
|||||
917 GTGACGAACCTTTAGTTTGCAGGCGATGAAGAAAGACGCTACAGAGCTT 966
|||||
298 LeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleTh 314
|||||
967 CTATTGGAGGTCATGAACACACCCGCCAGCAGCTCAACCTCATTGTCAT 1016
|||||
314 rTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluGluIleL 331
|||||
1017 GTTCTGGTCTGTAACACACAGAAAGTGGTGCAGAAAGTCAAGAGAGAGTTC 1066
|||||
331 ySerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMet 347
|||||
1067 AGGAGAAAGTTGAATGGGCATGTATACACCTGGAAGAGGCTTGAGTATG 1116
|||||
348 GluThrLeuGluGlnLeuLysTyrIleGlyCysValIleLysGluThrLe 364
|||||
1117 GAGCTGTTGGACAGCTGAAGTACACTGATGTGTGATTAAAGAGACTCT 1166
|||||
364 uArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysThrP 381
|||||
1167 TAGAATCAACCCCTCCTGTTCCCGAGGATTTCAGAGTCGACTCAAAACCT 1216
|||||
381 heGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSer 397
|||||
1217 TTGAATTGAATGTTTACCAATTCCTAAAGGATGGAAGTCATTACAGC 1266
|||||
398 IleCysaspThrHisAspValAlaAspIlePheThrAsnLysGluGluPh 414
|||||
1267 ATCTGTGACACGACGATGTGGCCGACGCTCTTCCAAACAAAGAGAGATT 1316
|||||
```

```
414 eAsnProAspArgPheIleValProHisProGluAspAlaSerArgPheS 431
|||||
1317 CCAGCCGGAGAGATTCATGAGCAAAAGTCTGAGAGACGGGTCCAGGTTTA 1366
|||||
431 erPheIleProPheGlyGlyGlyLeuArgSerCysValGlyLysGluPhe 447
|||||
1367 ACTACATCCCTTCGAGAGAGAGATCCAGGATGTGTGTGGCAAGAGTTC 1416
|||||
448 AlAlaLysIleLeuLysIlePheThrValGluLeuAlaArgHisCysAs 464
|||||
1417 GCCAAAGTGTACTCAAGATCTTTTGTAGTTAAGCAGCATTCGCAA 1466
|||||
464 pTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrValT 481
|||||
1467 TTGATTCCTCTCAAAAGGAGACCCCGACACATGAAGAAACAGGCCGACTATT 1516
|||||
481 yrProValAspAsnLeuProAlaArgPheThrTyrPhe 493
|||||
1517 ACCCAGTGACAAATCTCCCTACCAAAATTCAGTATGTTAT 1554
|||||
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-36

seq_documentation_block:

; Sequence 36, Application US/08882164D

; Patent No. 6306624

; GENERAL INFORMATION:

; APPLICANT: Petkovich, P. Martin, White, Jay A.,

; APPLICANT: Beckett, Barbara R., Jones, Glenville

; TITLE OF INVENTION: Retinoid Metabolizing Protein

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Blake, Cassels & Graydon

; STREET: Box 25, Commerce Court West

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5L 1A9

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage

; COMPUTER: COMPAQ, IBM PC compatible

; OPERATING SYSTEM: MS-DOS 5.1

; SOFTWARE: WORD PERFECT

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/882,164D

; FILING DATE: June 25, 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/667,546

; FILING DATE: June 21, 1996

; APPLICATION NUMBER: 08/724,466

; FILING DATE: October 1, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Hunt, John C.

; REGISTRATION NUMBER: 36,424

; REFERENCE/DOCKET NUMBER: 50767/00010

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 863-4344

; TELEFAX: (416) 863-2653

; INFORMATION FOR SEQ ID NO: 36:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2677 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-882-164D-36

alignment_scores:

Quality: 1251.00 length: 669

Ratio: 4.075 Gaps: 7

Percent Similarity: 45.889 Percent Identity: 43.797

alignment_block:

US-09-668-482-32 x US-08-882-164D-36 ..


```
288 ..... 288
2476 GTCTTCCTCCAGAACTCTCAGTTCGATTCGTAGTAATCCTTCTGTCAAA 2525
289 .....Ala..Leu 290
2526 CCGCAGGAGACTGTGTGAGAAATGTGGGTCTCACTCTATCTTAGGCACCTA 2575
291 LysGlnSerSerThrGluLeuLeuPheGlyGlyHisGluThrThrAlaSe 307
2576 AAGCAATCTTCAACCGAACTCCTTTGGAGACACGAAACCACGGCCAG 2625
307 rAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValIleug 324
2626 TGCAGCCACATCTCTGATCATTACTGGGGCTCTACCCACATGTTCTCC 2675
324 ln 324
2676 AG 2677
```

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-724-466B-11

seq_documentation_block:

```
; Sequence 11, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-724-466B-11
```

alignment_scores:

Quality:	604.00	Length:	117
Ratio:	5.207	Gaps:	0
Percent Similarity:	99.145	Percent Identity:	97.436

alignment_block:

US-09-668-482-32 x US-08-724-466B-11 ..

Align seg 1/1 to: US-08-724-466B-11 from: 1 to: 351

```
296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerIle 312
1 GAACTCCTCTTTGGAGGACACGAAACCACGGCCAGTGACGACATCTCT 50
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluG 329
51 GATCACTTACCTGGGGCTTACCACCATGTTCTCCAGAAAGTGCAGAGAG 100
329 IuIleLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 345
101 AGCTGAAGAGTAAGGTTTACTTTGCAAGAGCAATCAAGACAACAAGTTG 150
346 AspMetGluThrLeuGluGlnLeuLysTyrIleGlyCysValIleLysG1 362
151 GACATGGAATTTTGGACAACACTTAATATACATCGGGTGTATTAAAGA 200
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuL 379
201 GACCCTTCGACTGAATCCCCCAGTTCACAGGAGGGTTTCGGGTGCTCTGA 250
379 ysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIle 395
251 AGACTTTTGAATTAAATGATACACAGATTCGCCAAGGGCTGGAATGTTATC 300
396 TyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLysG1 412
301 TACAGTATCTGTGATCATCATGATGTGGCAGAGATCTTCACCAACAAGA 350
412 u 412
351 A 351
```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-882-164D-11

seq_documentation_block:

```
; Sequence 11, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-11

alignment_scores:
Quality: 604.00 Length: 117
Ratio: 5.207 Gaps: 0
Percent Similarity: 99.145 Percent Identity: 97.436

alignment_block:
US-09-668-482-32 x US-08-882-164D-11 ..

Align seg 1/1 to: US-08-882-164D-11 from: 1 to: 351

296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLe 312
1 GAACCTCCTCTTGAGAGGACACGAAACACGGCCAGTGCCAGCCACATCTCT 50
312 uUleThrTyLeuGlyLeuTyrrProHisValLeuGlnLysValArgGluG 329
51 GATCAGCTTACTGGGGCTCTACCCACATGTTCTCCAGAAAGTGCAGAG 100
329 LuLeuLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 345
101 AGCTGAAGAGTAAGGGTTTACTTTGCAAGAGCAATCAAGACAACAAGTTG 150
346 AspMetGluThrLeuGlnGlnLeuLysTyrIleGlyCysValIleLysG 362
151 GACATGGAAATTTTGAACAACCTTAATACATCGGGTGTGTATTATAGGA 200
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuL 379
201 GACCCCTTCGACTGAATCCCCAGATTCAGAGGGGTTTCGGGTGCTCTGA 250
379 ySThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIle 395
251 AGACTTTTGAATTAATGATACACAGATTCCTCCAAAGGGCTGGAATGTTATC 300
396 TyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLysG 412
301 TACAGTATCTGTGATACTCATGATGTGCGAGAGATCTTCACCAACAAGA 350
412 u 412
351 A 351

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-622-166A-1

seq_documentation_block:

; Sequence 1, Application US/08622166A
; Patent No. 5952545
; GENERAL INFORMATION:
; APPLICANT: KONCZ, CSABA
; APPLICANT: MATHUR, JAIDEEP
; APPLICANT: SZEKERES, MIKLOS
; APPLICANT: ALTMANN, THOMAS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
; TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,166A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 0147-0153P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: A. thaliana
; IMMEDIATE SOURCE:
; LIBRARY: lambda gt10
; CLONE: C204
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 48..1466
; US-08-622-166A-1

alignment_scores:
Quality: 450.00 Length: 503
Ratio: 1.562 Gaps: 15
Percent Similarity: 57.256 Percent Identity: 26.640

alignment_block:
US-09-668-482-32 x US-08-622-166A-1 ..

Align seg 1/1 to: US-08-622-166A-1 from: 1 to: 1608

6 LeuLeuAlaSerAlaLeuCysThrPheValLeuProLeuLeuPheLe 22
72 CTCCTCTCTTCATCGCCGCCGCTTCCTCCTACTC..... 110
22 uAlaAlaLeuLysLeuTrpAspLeuTyrCysValSerSerArgAspArg 39
111CGCCGTACAGCTT 123
39 erCysAlaLeuProLeuProProGlyThrMetGlyPheProPhePheGly 55
124 ACCGTGCGATGGGTCTGCCTCCGGGAAGCCTTGCTCTCCTCTGATAGA 173
56 GluThrLeuGlnMetVal.....LeuGlnArgArgLysPh 67
174 GAGACTTTTCAGCTGATCGGAGCTTACAAACAGAGAACCCCTGAGCCTTT 223
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 84
224 CATCGACGAGAGAGTAGCCCGGTACGGTTCGGTTTTCATGACGACATCTT 273
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
274 TTGGTGAACCGACGATTTTCTACGTGACCCGGAAACGAAACCGGTTGTT 323
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
324 CTTTCAGAACGAAAGGAGCTTTTGAAGTGTCTTATCCCTGCTTCATTTG 373
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
374 TAACCTTTTGGGGAACACATCTCTCTCTTATGAAAGGTTCTTTGCATA 423


```
80 sthrHisLeupheGlyArgProThrValArgValMetGlyAlaAspAsnV 97
   ::::: ||| ||| |||
295 CCTCAAGATGGATCTTACACATGTTGCCGCTTCCACACCCGACATGG 344
   ::::: ||| ||| |||
97 aLArgArgIleLeuLeuGlyGluHisArgLeuValSerValHisTrpPro 113
   ::::: ||| ||| |||
345 CTCGCCAAGTCTTCAAGTCCAGACACACATCTTCTCG...AACCGGCCA 391
   ::::: ||| ||| |||
114 AlaserValArgThrIleLeuGlyAlaGlyCysLeuSerAsnLeu...H1 129
   ::::: ||| ||| |||
392 GCCACCATA.....GCCATCAGCTACCTCACCCTA 420
   ::::: ||| ||| |||
129 sAspSerSer.....HisLysG 135
   ::::: ||| ||| |||
421 TGACCGAGCCGACATGGCTTCGCTACACGCGCTTTTGGCGTCAGA 470
   ::::: ||| ||| |||
135 InArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGlnCys 151
   ::::: ||| ||| |||
471 TGGGTAAACTCTGCGTCAATAATATTAGCCGGAACGACCGAGTGC 520
   ::::: ||| ||| |||
152 Tyr.....ValleuValIleAlaG1 158
   ::::: ||| ||| |||
521 TGGGAGTCCGTCGAGACGAGTCCGACTCGGCGAGTACGAGTGGTCCGTC 570
   ::::: ||| ||| |||
158 uGluValSerSerCysLeuGluGlnTrpLeuSerCysGlyGluArgGlyL 175
   ::::: ||| ||| |||
571 CAATATTGGGTCGACG.....GTGAATATCCGCGAG..... 601
   ::::: ||| ||| |||
175 euLeuValTyrProGluValLysArgLeuMetPheArgIleAlaMetArg 191
   ::::: ||| ||| |||
602 ..CTGGTTTTCCTCTGACGAAGAATATTACTTACAGGCGGCTTTGGG 649
   ::::: ||| ||| |||
192 IleLeuLeuGlyCysGluProGlyProAlaGlyGlyGluAspGluG1 208
   ::::: ||| ||| |||
650 ACGATC.....TCGCATGAGGACGACGA 672
   ::::: ||| ||| |||
208 nGlnLeuValGluAlaPheGluGluMetThrArg.....AsnLeuPheS 223
   ::::: ||| ||| |||
673 CGAGTTCGTGGCCATCTGCAAGAGTTTTCGACGCTGTTGGTCTTTTA 722
   ::::: ||| ||| |||
223 erLeuProIleAspValProPhe...SerGlyLeuTyrArgGlyVal... 237
   ::::: ||| ||| |||
723 ATATAGCTGATTATTCCTTGGCTCAAAATGGTTCCTCAGGGGATTAAAC 772
   ::::: ||| ||| |||
238 .....LysAlaArgAsnLeuIleHisAlaArgIleGluGluAs 250
   ::::: ||| ||| |||
773 GTCAGGCTCAACAAGGCGAGGCGGCTTGATGGGTTTATTGACAAGAT 822
   ::::: ||| ||| |||
250 nIleArgAlaLysIleArgArg....LeuGlnAlaThrGluProAspG 265
   ::::: ||| ||| |||
823 CATCGACGATCATATACAGAAGGGGAGTAAAAAATCGGAGGAGGTGATA 872
   ::::: ||| ||| |||
265 lYgIlyCysLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArg 281
   ::::: ||| ||| |||
873 CTGATATGGTAGATGATTACTTCTTTTACGGTGAGGAAGCCAAAGTA 922
   ::::: ||| ||| |||
282 GlyGluArgLeuAspMetGln...AlaLeuLys..... 291
   ::::: ||| ||| |||
923 AGCGAATCTGACGATCTTCAAAATTCATCAAACTCACCAGAACACACAT 972
   ::::: ||| ||| |||
292 .GlnSerSerThrGluLeuLeuPheGlyGlyHisGluThrAlaSerA 308
   ::::: ||| ||| |||
973 CAAGCTATCATGACGTAATGTTGGAGGAGCGGAACGGTGCGCTCCG 1022
   ::::: ||| ||| |||
308 laAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuGln 324
   ::::: ||| ||| |||
1023 CGATTGAATGGCCATGACGAGCTGATGAAAAGCCAGAAAGATCTAAAG 1072
   ::::: ||| ||| |||
325 LysValArgGluGluIleLysSerLysGlyLeuLeuCysLysSerAsnG1 341
   ::::: ||| ||| |||
1073 AAGTCCCAACAGAA.....CTGCCGTGCTGCTGGTCT 1107
   ::::: ||| ||| |||
341 nAspAsnLysLeuAspMetGluThrLeuGluGlnLeuLysTyrIleGlyC 358
```

```
1108 TGACCGCGAGTCGAAGAGAAAGACTTCGAGAAGCTCACCTACTTGAAT 1157
   ::::: ||| ||| |||
358 ySValIleLysGluThrLeuArgLeuAsnProProValProGlyGlyPhe 374
   ::::: ||| ||| |||
1158 GCGTACTGAAGGAAGTCTTCGCCCTCCACCACCCACTCCCTCCTC 1207
   ::::: ||| ||| |||
375 ArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysG1 391
   ::::: ||| ||| |||
1208 CACGAGACTGCCGAGACGCCGAGGTCGGCGCTACTACATTCGGCGAA 1257
   ::::: ||| ||| |||
391 yTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaAspIleP 408
   ::::: ||| ||| |||
1258 ATCGCGGCTGATGATCAACGCGTCCGATCGCGCGGACAGAACTCGT 1307
   ::::: ||| ||| |||
408 heThrAsnLysGluGluPheAsnProAspArgPheIleValProHisPro 424
   ::::: ||| ||| |||
1308 GGGCCGACCCAGATACGTTAGGCCCTCCAGGTTCTCAAGACGGTGTC 1357
   ::::: ||| ||| |||
425 GluAsp.....AlaserArgPheSerPheIleProPheGlyGlyLe 439
   ::::: ||| ||| |||
1358 CCCGATTTCAAAGGAGAACACTTCGAGTTCATCCCATTCGGGTCAGGTCG 1407
   ::::: ||| ||| |||
439 uArgSerCysValGlyLysGluPheAlaLysIleLeuLeuLysIlePheT 456
   ::::: ||| ||| |||
1408 TCGGCTTGCCCCGGTATGCACTCGGACTCTACGCGGCTAGAGACG...A 1454
   ::::: ||| ||| |||
456 hrValGluLeuAlaArgHisCys...AspTrpGlnLeuLeuAsnGlyPro 471
   ::::: ||| ||| |||
1455 CTGTGGCTACCTCTCTTCACTGTTTCACTGAGGAGTTG.....CCG 1495
   ::::: ||| ||| |||
472 ProThrMetLysThrSer 477
   ::::: ||| ||| |||
1496 GACGGGATGAACCGAGT 1513
   ::::: ||| ||| |||

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-09-091-432-1

seq_documentation_block:
: Sequence 1, Application US/09091432
: Patent No. 5981837
: GENERAL INFORMATION:
: APPLICANT: Chapelle, Clint
: TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition
: FILE REFERENCE: 7024-325
: CURRENT APPLICATION NUMBER: US/09/091,432
: CURRENT FILING DATE: 1998-06-18
: EARLIER APPLICATION NUMBER: PCT/US96/20094
: EARLIER FILING DATE: 1996-12-19
: EARLIER APPLICATION NUMBER: US 60/009,119
: EARLIER FILING DATE: 1995-12-22
: EARLIER APPLICATION NUMBER: US 60/013,388
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: Microsoft word 2.0C
: SEQ ID NO 1
: LENGTH: 1838
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
: US-09-091-432-1

alignment_scores:
Quality: 337.00 Length: 515
Ratio: 1.221 Gaps: 23
Percent Similarity: 53.592 Percent Identity: 25.825

alignment_block:
US-09-668-482-32 x US-09-091-432-1 ..
Align seg 1/1 to: US-09-091-432-1 from: 1 to: 1838
15 ValleuProLeuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTy 31
||||: |||:|||||:||||:|
```

```
80 GTTGTCTCTCTTTTCATCTCATCAGC..... 106
31 rcysvalserSerArgaspargSerCysAlaLeuProLeuProProGlyT 48
   ::::: ||| ||| ||| ||| ||| |||
107 .TTCATTCACACGGCGCGAAGCCT.....CCATATCCTCCCGGTC 146
48 hrmetGlyPheProPhePheGlyGluThrLeuGlnMetValLeuGlnArg 64
   ||:::||||:::||||::: |||::|||
147 CACGAGGTTGGCCCATCATAGCACAATGTTAATGATGACCAACTCACC 196
65 ArgLysPheLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysTh 81
   ::::: ||| ::::: ||| ||| ||| |||
197 CACCGTGGTTTAGCCAATTAGCTAAAAAGTATGGCGGATG.....TG 240
81 rHisLeu...PheGlyArgProThrValArgValMetGlyAlaAspAsnV 97
   ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
241 CCATCTCCGATGGGATTCCTCCATATGTACGCTGTCTCATCACCAGG 290
97 aArgArgIleLeuLeuGlyGluHisArgLeuValSerValHisTrpPro 113
   || ||| ::::: ||| ::::: ||| ::::: |||
291 TGGCTCGACAAGTCCCTTCAAGTCCAGACAGCGCTTCTCGAACCGGCT 340
114 AlaSerValArgThrIleLeuGlyAlaGlyCysLeuSerAsnLeu...Hi 129
   ||::: ||| ::::: ||| ||| ||| |||
341 GCAACTATA.....GCTATAGCTATGACTTGA 369
129 sasPserSer.....HisLysG 135
   ::::: ||| ::::: ||| ::::: |||
370 CGACCGAGCGGACATGGCTTTCGCTACACGACCGTTTGGAGACAGA 419
135 lnArgLysLysValIleMetGlnAlaPheSerArgLysAlaLeuGlnCys 151
   ::||| ||| ::::: ||| ||| ||| |||
420 TGAGAAAAAGTGTGTCTCATGAAGGTGTTAGCCGTAAGAGAGCTGAGTCA 469
152 TyrValLeuVal.....IleAlaGluGluValSerSe 162
   ::::: ||| ::::: ||| ||| ||| |||
470 TGGGCTTCAGTTCGTGATGAAGTGACAAATGGTCCGGTCGGTCTCTTG 519
162 rCysLeuGluGlnTrpLeuSerCysGlyGluArgGlyLeuLeuValTyrP 179
   ::::: ||| ::::: ||| ||| ||| |||
520 TAAGTTGGTAAGCCTATAAACGTGGGGAGCAA.....ATTTTTG 560
179 roGluValLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGly 195
   ::::: ||| ::::: ||| ||| ||| |||
561 CACTGACCCGCAACATACTTACCGGCGAGCGTTGGG.....TCAGCC 604
196 CysGluProGlyProAlaGlyGlyGlyLysAspGluGlnLeuValG1 212
   ||||| ||| ::::: ||| ||| ||| |||
605 TGGGAGAAGGA.....CAAGACGAGTTCATAG 633
212 uAlaPheGluGluMetThrArg.....AsnLeuPheSerLeuProIleA 227
   ::::: ||| ::::: ||| ||| ||| |||
634 AATCTTACAAGAGTCTCTAAGCTTTTGGAGCCTTCAACGTAGCGGAT 683
227 spValProPheSerGlyLeuTyr.....ArgGly 236
   ::||| ::::: ||| ||| ||| |||
684 TCATACCATATTTCCGGGTGGATCGATCCGCAAGGATAAACAAGCGGCTC 733
237 ValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsnIle.... 251
   ||||| ||| ||| ::::: ||| ||| ||| |||
734 GTGAAGGCCCGTAATGATCTAGACGATTATGTACGATATATTCGATGA 783
252 .ArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGly... 266
   ::::: ||| ::::: ||| ||| ||| |||
784 ACATATGAAGAAGAGAGAAATCAAAACGCTGTGATGATGGGATGTTG 833
267 .....CysLysAspAlaLeuGlnLeuLeuIleGluHisSer 278
   ||||| ||| ||| ::::: ||| ||| ||| |||
834 TCGATACCGATATGTTGATGATCTTCTTGTCTTTTACAGTGAAGAGGCC 883
279 TrpGluArgGlyGluArgLeuAspMetGln..... 288
   ::||| ||| ||| ||| ||| |||
884 AAATTAGTCAAGAGACAGCGGATCTCAAAATTCATCAAACTTACCCG 933
```

```
289 ....AlaLeuLysGlnSerSerThrGluLeuLeuPheGlyGlyHisGluT 304
   ::||| ::::: ||| ||| ||| ||| |||
934 TGACAATATCAAAACAATCATCATGACGCTTATGTTGGAGGAACGAAA 983
304 hrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrPro 320
   ||::: ||| ||| ::::: ||| ||| ||| |||
984 CGGTAGCGTGGCGATAGAGTGGCCCTTAACGAGTATATTACGAGCCCC 1033
321 HisValLeuGlnLysValArgGluGluIle...LysSerLysGlyLeu 336
   ::::: ||| ::::: ||| ||| ||| |||
1034 GAGGATCTAAACGGGTCCACAAGAACTCGCCGAAGTCGTTGACTT... 1081
336 uCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeuGluGlnL 353
   ||::: ||| ::::: ||| ||| ||| |||
1082 .....GACAGACGAGTTGAAGAAATCCGACATCGAGAACT 1115
353 euLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAsnProPro 369
   || ||| ::::: ||| ||| ||| ||| |||
1116 TGACTTATCTCAATGACACACTCAAGAAACCTTAAGATGACACCACCG 1165
370 ValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyr 386
   ::||| ::::: ||| ::::: ||| ||| ||| |||
1166 ATCCCTCTCTCTCCACGAAACCGCGAGGACACTAGTATCGACGGTTT 1215
386 rGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAspThrHisA 403
   ::||| ||| ::::: ||| ||| ||| |||
1216 CTTCATTTCCCAAGAAATCTGTGTGATGATCAACGCGTTTGCATAGAC 1265
403 spValAlaAspIlePheThrAsnLysGluGluPheAsnProAspArgPhe 419
   ::||| ::::: ||| ||| ||| ||| |||
1266 GCGACCCAACTCTTGACGTACCCGACACAGCTTTAGACCATCGAGGTTT 1315
1316 TTGGAACCGGGCGGTACCGGATTTCAAGGAGCAATTCGAGTTATACC 1365
434 oPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleL 451
   ||||| ::::: ||| ||| ||| ||| |||
1366 GTTCGGGTGGGTGCTAGATCGTCCCGGGTATGCAA..... 1402
451 euLeuLysIlePheThrValGluLeuAla.....ArgHisCys 463
   ||| ::::: ||| ||| ||| ||| |||
1403 ..CTAGGTTATACGCGCTTGACTTAGCCGTGGCTCATATATTACATTGC 1450
464 ...AspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSer 477
   ||| ::::: ||| ||| ||| ||| |||
1451 TTCACGTGGAAATTA.....CCTGATGGGATGAACCAAGT 1486
seq_name: /cgn2_6/plodata/2/ina/6B_COMB.seq:US-09-103-840A-2
seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103, 840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
```

OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

alignment_scores: Quality: 329.00 Length: 456
 Ratio: 1.348 Gaps: 12
Percent Similarity: 53.509 Percent Identity: 24.123

alignment_block:

US-09-668-482-32 x US-09-103-840A-2/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

```
32 CysValSerSerArgAspArgSerCysAlaLeuProLeuProProGlyTh 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120837TGCCTGTCGGCGGAGACGCGCGGATGCTTGGCGCGCGCCACCGGCT 4120788

48 rMetGlyPheProPhePheGlyGlyThrLeuGlnMetValLeuGln... 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120787TGCCTGCGCGCGCGGATGCTTGGCGCGCGCGCGCTACCGCGGT 4120738

64 .....Arg.ArgGlyPheLeuGlnMetLysArgArgLy 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120737GCCTACCTGCGCGGCGGAGCGCGGATGCTGCGACTGATCCGTCGG.. 4120690

74 sTyrglyPheIle..... 78
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120689.TACGACCGCATGACGATGCCCATTTCTCAGCCTGGGTGACGTTGCGA 4120641

79 .....TyrLysThrHisLeuPheGlyArgProThrVal 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120640TCGCTGTCGACTGCGCGGCGGAGCGGATGCTTTCACCGCTCCACCGAC 4120591

90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyGlnHisAr 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120590GTCCTGCTCGGGGCGGAGGGGTG..... 4120567

106 gleuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAlaG 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120566.....GGCGCGGCGACG.....GCGATCTATGGGTGCG 4120539

123 LysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLysVal 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120538GTCGATGTTTGTCCAGAGGAGCGCGGACATCTTCGACGCGAAACTC 4120489

140 IleMetGlnAlaPheSerArgGluAlaLeuGlnCysTyrValLeuValIl 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120488TTGACGCGCGCATATACATGGCGCTGCGTGCATGTTACGTTCCGATTAT 4120439

156 eAlaGluGluValSerSerCysLeuGluGlnTrpLeuSerCysGlyGluA 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120438CGAAGACTCCACTCGCGCGCGCATGATACATGGCGGCTT.....GACC 4120395

173 rgGlyLeuLeuValTyrProGluValLysArgLeuMetPheArgIleAla 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120394GTCGCTTTCGATGTTGACGGGTGGCGGCTCGCTGATGCTGATGATC 4120345

190 MetArgIleLeuLeuGlyCysGluProGlyProAlaGlyGlyGlyGluAs 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120344GTCAAGTCAATATTCGGGGTGCAC.....GACCCGAGGA 4120310

206 pGluGlnGlnLeuValGluAlaPheGluGluMetThrArgAsnLeuPheS 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120309GGTCCGGCTTGGCGCGCGCGGCTTCGAGCGGCTGTGAATCTTGTGTTT 4120260

223 erLeuProIleAspValProPheSer.....GlyLeuTyrArg 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120259CAGAGCAGTTGACCGTCCGCTACGCACTGCGTGGTTGGCGCACTACGA 4120210

236 GlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsnIleAr 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120209GTCCTGCGCTGCACGCGCGCGCGCAACACCGAGATGACGACGCTGTCAT 4120160
```

```
252 gAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCysLysA 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120159GGCGCTCATTTGCGGAGCGCGCGCGGATCCACCGCTGGCGGAGACACG 4120110

269 sPaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGluArgLeu 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120109ACGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4120060

286 AspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGlyGlyHi 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120059TCAGACAGCAGATCCGCGGAGCATCTGATCACCTGCTGCTGCTGCTG 4120010

302 sGluThrThrAlaSerAlaAlaThrSerLeuIleThrThrLeuGlyLeuT 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4120009TGAACACAGCGGACACGATGCGGCTGGCGGCTTCGACTTCTGCTGCACC 4119960

319 yrProHisValLeuGlnLysValArgGluGluIleLysSerLysGlyLeu 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4119959ATCCGAGCGCTTGGCGCGCGCTGCGAGCGGAGCGCGCTGGCGTGGG... 4119913

336 LeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeuGluG 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4119913..... 4119913

352 nLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAsnProp 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4119912.GAAGCTTTCACGACGCGCGGTGATCAACGAGACGTGCGGCTCCGCC 4119864

369 rovalProGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGly 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4119863CCGCGCGGTGACGGCTCGTGTGCGCGCGCAACACCTAACCATCGCGGT 4119814

386 TyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAspThrHi 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4119813TACCGTGTGAGGCTGCGACACAGATCGTGTCCACATCATCGCATCAA 4119764

402 sAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAspArgP 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4119763CCGACGCGGAGGTGATGACGACACACACAGAGTCCGCGCGGAGCGGT 4119714

419 heIleValProHisProGluAspAlaSerArgPheSerPheIleProPhe 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4119713TCTCTCGGTACGCGCGCGCA.....ACCTATGCTGCTGCTGCTGCT 4119673

436 GlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLe 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4119672GGCGCGCGGTGAACGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 4119629

452 uLysIlePheThrVal 457
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4119628AGAGTTGATCACCGTA 4119613

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1

seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103, 840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
```


ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

alignment_scores:
Quality: 329.00 Length: 456
Ratio: 1.348 Gaps: 12
Percent Similarity: 53.509 Percent Identity: 24.123

alignment_block:
US-09-668-482-32 x US-09-103-840A-1/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529

```
32 CysValSerSerArgAspArgSerCysAlaLeuProLeuProGlyTh 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128642TGCGGTTCCGGAGACGCGCGGATGCTTGGGGCGCCACCGGCT 4128593

48 rMetGlyPheProPhePheGlyJutThrLeuGlnMetValLeuGln... 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128592TGCCGGCCCGCTGATTTCGCGCGGCTTCGCGGCCCTACCGGTT 4128543

64 .....Arg.ArgLysPheLeuGlnMetLysArgArgLy 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128542GCCTACCTGGCGGGGTGGCGAGCGCGGATGCTGCGACTGATCCGTCGG.. 4128495

74 sTyrgLysPheLeu..... 78
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128494.TACGAGCCGATCATGACGATGCCCATTCACGCTGGGTGACGTTGCCGA 4128446

79 .....TyrLysThrHisLeuPheGlyArgProThrVal 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128445TCGTGTCGACTCGGCGCTGGCCAGAGAGGTTTTCACCGCTCCACCGAC 4128396

90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyLuhHisAr 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128395GTCTGCTGGGGGGCGAGGGGGTG..... 4128372

106 gLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAlaG 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128371.....GGCGCGGACGG.....GCGATCTATGGGTCCGG 4128344

123 lYcysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLysVal 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128343GGTCGATGTTGTCCAGAGAGAGCGGACATCTTCGACGTGAAACATC 4128294

140 lLeuMetGlnAlaPheSerArgGluAlaLeuGlnCysTyrValLeuValIl 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128293TTGACGCGCGCCATTACATGGCGCTGCGTGCATGCTTACGTTCCGATTAT 4128244

156 eAlaGluGluValSerSerCysLeuGlnIleTrpLeuSerCysGlyGluA 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128243CGAGAACTCCACTCGCGCGGCGATGATGACCGCGGT....GACC 4128200

173 rgGlyLeuLeuValTyrProGluValLysArgLeuMetPheArgIleAla 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128199GTCGTTGGATGTTGACGGGTGGCGCGGCTGCTGATGCTGATGTCATC 4128150

190 MetArgIleLeuLeuGlyCysGluProGlyProAlaGlyGlyGlyLys 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128149GTCAAGTCAATTCGGGGTGCAC.....GACCCCGAGGA 4128115

206 pGluGlnGlnLeuValGluAlaPheGluGluMetThrArgAsnLeuPheS 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128114GGTCCGGCGCTTGGCGCGCGCTTCGAGCGGCTGTGAATCTTGTGTTT 4128065

223 erLeuProIleAspValProPheSer.....GlyLeuTyrArg 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4128064CAGAGCAAGTTGACCGTGGGCTACGCACTGCGGTGGCGCACTACGA 4128015

236 GlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsnIleAr 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
4128014GTCTGGCCTGCAAGCGCGCCGCCAACACCGAGATCGACGACCTGTCAT 4127965

252 gAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCysLysA 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127964GGCCTCATTTGGCCAGCGCGCGCATCCACCGCTGGCGAGCGACAGC 4127915

269 sPAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluArgLeu 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127914ACGTCTGTGCTGCTGAGCGCGCGGAGAGATCCGGGAACAGTTG 4127865

286 AspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheGlyGlyHi 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127864TCAGACAGCGAGATCCGCGACGATCTGATCACCCCTGCTGCGGGGCA 4127815

302 sGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyLeuT 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127814TGAACACAGCGCGACACGTTGGCGTGGCGTTGCACTTGCTGCTGCACC 4127765

319 yRProHisValLeuGlnLysValArgGluGluIleLysSerLysGlyLeu 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127764ATCCGAGACGCTTGGCGCGCGTGCAGAGCGAGCGCTGGCGGTGG... 4127718

336 LeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeuGluGl 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127718..... 4127718

352 nLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAsnProP 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127717.GAAGCCTTCACGACGCGCGGTGATCAACGAGACGTTGCGGTGCTCCGC 4127669

369 roValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGly 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127668CCGCGCGCTGACGCGCTGCTGCGCGCGCACACCACTAACCATCGCGGT 4127619

386 TyrGlnIleProLysGlyTyrAsnValIleTyrSerIleCysAspThrHi 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127618TACGCTGTGGAGGCTGGCACGAAATCGTGTCCACATCATCGCATCAA 4127569

402 sAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAspArgP 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127568CCGACCGCGAGGTGTATGACACCCACACGAGTTCCGCCGAGCGGT 4127519

419 helleValProHisProGluAspAlaSerArgPheSerPheIleProPhe 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127518TCTCTCGGTACGCGCGCGCAA.....ACCTATGCTTGGGTCCGTTCC 4127478

436 GlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLe 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127477GGCGCGCGGTGAACGCTGTCTGGTGGCGAATTTTCG.....ATGCG 4127434

452 uLysIlePheThrVal 457
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4127433AGAGTTGATCACCGTA 4127418

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103, 840A
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
```


224	GAGAC.CGGCTGGTGTGCGTCCACTGGCCAGCGTCGGTGGCACCATTCT	272
120	UGLYAlAGLYCysLeuSerAsnLeuHisaspSerSerHisLysGlnArgL	137
273	GGGATCTGGCTGCCTCTCTTAACCTGCACGACTCCTCGCACAAAGCAGCA	322
137	ysLysValIleMetGlnAlaPheSerArgGlnAlaLeuGlnCysTyrVal	153
323	AGAAGTGATTTATGCGGGCCTTCAGCCGCGAGGA.CTCGAATGCTACGTG	371
154	LeuValIleAlaGlnGluValSerSerCysLeuGlnGlnTyrLeuSerCys	170
372	CCGGTGATCACCCGAGGAAGTGGCGACAGACCTTGAGCAGTGGCTGAGCTG	421
170	scLysGluArgGlyLeuLeuValTyrProGluValLysArgLeuMetPheA	187
422	CGGGAGCGCGGGCCCTCCTGGTCTACCCCGAGGTGAAGCGCCTCATGTTCC	471
187	rgIleAlaMetArgIleLeuLeuGlnCysGluProGlyProAlaGlyGly	203
472	GAATCGCCATGCGCATCCTACTGGGCTGCGAACCCCAACTGGCGGGCGAC	521
204	GlyGluAspGluGlnGlnLeuValGlnAlaPheGluGluMetThrArgAs	220
522	GGGGACTCCGAGCAGCAGCTTGAGAGCCCTTCGAGGAATGACCCGCAA	571
220	nLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyV	237
572	TCTCTTCTGCTGCCCATCGACGTGCCCTTCAGCGGGCTGTACCGGGGCA	621
237	alYsAlaArgAsnLeuIleHisAlaArgIleGluGlnAsnIleArgAla	253
622	TGAAGCGCGGAACCTCATTCACGCGCGCATCGAGCAGAACAATTCGGCC	671
254	LysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCysLysAspAl	270
672	AAGATCTCGGGCTGCGGGCATCCGAGGCGGCGCAGGGCTGCAAGACGC	721
270	aleuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGluArgLeuAspm	287
722	GCTGCAGCTGTGATCGACACCTCGTGGAGAGGGGAGAGCGGCTGGACA	771
287	etGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGlyGlyHisGlu	303
772	TGCAGGCACTAAAGCAATCTTCAACCGAACTCCTTTGGAGGACACGAA	821
304	ThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrPr	320
822	ACCACGGCCAGTGCAGCACATCTGTATCACTTACCCTGGGGCTTACCC	871
320	ohisValLeuGlnLysValArgGluGluIleLysSerLysGlyLeuLeu	337
872	ACATGTCTCCAGAAAGTGGGAGAAGAGCTGAAGAGTAAAGGTTTACTTT	921
337	ysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeuGlnGlnLeu	353
922	GCAAGAGCAATCAAGACACAAGCTTGGACATGGAATTTTGGAACTTA	971

seq_documentation_block:			
LOCUS	AL532445	893 bp	mRNA
DEFINITION	AL532445 LTI_NFL001_NBC4 Homo sapiens CDNA clone CS0DM014YL13	EST	13-FEB-2001
ACCESSION	AL532445	prime, mRNA	sequence.
VERSION	AL532445.1	GI:12795938	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 893)		
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.		

```

TITLE
JOURNAL
COMMENT
FEATURES
source
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
location/Qualifiers
1.893
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODM014YL13"
/clone_11b="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT
196 a 253 c 275 g 167 t 2 others
ORIGIN

```

	alignment_scores:	
Quality:	1255.00	Length: 307
Ratio:	4.498	Gaps: 4
Percent Similarity:	90.879	Percent Identity: 86.319

alignment_block:

Align seg 1/1 to: AL532445 from: 1 to: 893

73	ArgLysI ^T YrGlyPheIle ^T YrLysThrHisLeuPheGlyArgPro Th Val	89
2	AGGAATACGGCTTCATCTACAAAGACGCACTGTTCGGGGCCCACT	51
89	LArgValMetClyAlaAspAsnValArgArgIleLeuLeuGlyL ^U HisA	106
52	ACGGGTGATGGCGCGGACAAATGTGGCGGCATCTTGCTCGGAGA.SACC	100
106	rgLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAla	122
101	GGCTGGTTCGGTCCACTGGCCAGCGTCGGTGCACCACTTCTGGGATCT	150
123	GlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArg.LysLysV	139
151	GGCTGCTCTTACTACCTGCACGACTCTCGACAAGCAGCGCAGA...	196
139	allMetGlnAlaPheSerArgGlnAlaLeuGlnCys ^T YrValLeuVal	155
197AGCCGGCAGGCACCTCGAATGCTACGTGCCGGTG	229
156	IleAlaGlnGluValSerSerCysLeuGlnGlnTrpLeuSerCysGlyG	172
230	ATCACCGAAGAGTGGGACGACGACTGGAGCAGTGCGTGAAGTGGCGGA	279
172	uArgGlyLeuLeuVal ^T YrProGluValLysArgLeuMetPheArgIleA	189
280	GGCGGGCTCCTGTGTACCCGAGGTGAAGCGCCTCATGTCCGAATCG	329
189	IaMetArgIleLeuLeuGlyCysGlnProGlyProAlaGlyGlyGlyGlu	205
330	CCATGCGCATCTACTGGGCTGGGAACCCCACTGCGCGGCACGGGGAC	379
206	AspGlnGlnLeuValGlnAlaPheGlnGlnMetThrArgAsnLeuPh	222

380 TCCGAGCAGCAGCTTGTGGAGGCCCTTCAGGAATGACCCGCAATCTCT 429
222 eSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyValLysA 239
|||||
430 CTCGCTGCCCATCGACGTGCCCTTCAGCGGGCTGTACCGGGGCATGAAG 479
239 laArgAsnLeuIleHisAlaArgIleGluGluAsnIleArgAlaLysIle 255
|||||
480 CGCGGAACCTCATTCACGCGCGCATCGAGACAACATTCGCCCAAGATC 529
256 ArgArgLeuGlnAlaThrGluProAspGlyGlyCysLysAspAlaLeuG 272
|||||
530 TCGGGCTGCGGCATCCGAGCGGGCCAGGCTGCMAAGAGCGCTGCA 579
272 nLeuLeuIleGluHisSerTrpGluArgGlyGluArgLeuAspMetGlnA 289
|||||
580 GCTGTGATCGAGCACTCGTGGAGAGGGGAGAGCGGCTGGACATGCAGG 629
289 laLeuLysGlnSerSerThrGluLeuLeu.PheGlyGlyHisGluThrTh 305
|||||
630 CACTAAGCAATCTTCACACCGAAGCTCTTTGGAGACACAGAAACCAC 679
305 rAlaSerAlaIaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisV 322
|||||
680 GGCAGTGCAGCCACATCTCTGATCACTTACCTGGGGCTTACCCCATG 729
322 alLeuGlnLysValArgGluGluIleLysSerLysGlyLeuLeuCysLys 338
|||||
730 TTCTCAGAAAGTGCAGAGAAGAGTGAGAGTAAGGGTTACTTTGCCAAG 779
339 SerAsnGlnAspAsnLysLeuAspMetGluThrLeuGlnLeuLysTyr 355
|||||
780 AGCAATCAAGACACAAGATTGGACATGAAATTTTGAACA.CTTAAATA 828
355 rIleGlyCysValIleLeuGluThrLeuArgLeuAsnProProValProg 372
|||||
829 CATCGGTGTGTATTAAAGAGACCTTCGACTGAATCCCCAGT...TCAG 875
372 lyGlyPheArgValAla 377
|||||
876 GAGGGTTTCGGGTGCT 892
seq_name: gb_est29:AL532444
seq_documentation_block:
LOCUS AL532444 1028 bp mRNA EST 13-FEB-2001
DEFINITION AL532444 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM014YL13 3
prime, mRNA sequence.
ACCESSION AL532444
VERSION AL532444.1 GI:12795937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1028)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..1028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DM014YL13"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCMVSPORT 6
vector. Library was normalized. Library was constructed
by life Technologies. Contact : Feng Liang life
Technologies, a division of invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 265 a 244 c 225 g 282 t 12 others
ORIGIN
alignment_scores:
Quality: 1237.00 Length: 282
Ratio: 4.739 Gaps: 3
Percent Similarity: 92.553 Percent Identity: 86.170
alignment_block:
US-09-668-482-32 x AL532444/rev ..
Align seg 1/1 to reverse of: AL532444 from: 1 to: 1028
217 MetThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLe 233
|||||
1005 ATGACCCGA...ATCTCTCGCTCCCATCGACTGCCCT...TCAGCGGC 962
233 uTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluA 250
|||||
961 TGTAACCGGCATGTAGGCGCAACCTTAATTAAACGCGGCATCGAGCAGA 912
250 snIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGly 266
|||||
911 AMATACGCCCAAGATCTGCGGGCTGCGGCATCCGAGCGGGCCAGGGC 862
267 Cys.LysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyG 283
|||||
861 TGCAAAGAAGACCGCGCTGCAGCTGTGATCGAGCACTCGTGGAAGAGGAG 812
283 luArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPhe 299
|||||
811 AGCGGTGACATGACGACACTTAAGCAATCTTCACCGCAACTCCTCTTT 762
300 GlyGlyHisGluThrThrAlaSerAlaIaThrSerLeuIleThrTyrLe 316
761 GGAGGACACGAAACACGCGCAGTGCAGCACATCTTGATMACTTACCT 712
316 uGlyLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerL 333
|||||
711 GGGGCTTACCCACATGTTCTCCAGAAAGTGCAGAGAAGCTGAAGAGTA 662
333 ySGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThr 349
|||||
661 AGGGTTTACTTTTGCAAGAGCAATCAAGACAACAAGTTGACATGGAAT 612
350 LeuGlnGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLe 366
611 TTGGAACAACCTTAATATACATCGGGTGTATTATTAAAGAGACCTTCGACT 562
366 uAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGluL 383
561 GAATCCCCCAGTTCAGAGAGGGTTTCGGGTTGCTCTGAAGACTTTTGAAT 512
383 euAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCys 399
511 TAAATGGAATACAGATTCCCAAGGCTGGAATGTTATCTACATATCTGT 462
400 AspThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnPr 416
461 GATACCTCATGATGTGGCAGAGATCTTCAACCAAGAAGAAATTTAATCC 412
416 oAspArgPheIleValProHisProGluAspAlaSerArgPheSerPheI 433
|||||

411 TGACCGATTCAAGTCGCTCACCAGAGATGCATCCAGCTTCAGCCTCA 362
433 leProPheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLys 449
|||||
361 TTCATTTGGAGAGCCTTAGAGCTGTAGGCAAGAATTGGCAAAA 312
450 lleLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpG1 466
|||||
311 ATTCTTCTCAAAATATTACAGTGCAGCTGSCCAGGCATTGTSACTGGCM 262
466 nLeuLeuAsnGlyProProThrMetLysThrSerProThrValTyrProV 483
|||||
261 GCTCTAAATGGACTTCCTACATGAAMWCYAGTCCCAACCGGTATCCTG 212
483 aLasPasnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||
211 TGGACAATCTCCTGCAAGATTCAACCATTTCCATGGGGAATC 168
seq_name: gb_est29:AL539667

seq_documentation_block:
LOCUS AL539667 847 bp mRNA EST 16-FEB-2001
DEFINITION AL539667 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF036YG04 3
prime, mRNA sequence.
ACCESSION AL539667
VERSION AL539667.1 GI:12869095
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 847)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. 847
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF036YG04"
/clone_lib="LTI_FL013_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dt) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 197 a 190 c 204 g 229 t 27 others
ORIGIN

alignment_scores:
Quality: 1001.00 Length: 231
Ratio: 4.859 Gaps: 0
Percent Similarity: 89.177 Percent Identity: 82.684

alignment_block:

US-09-668-482-32 x AL539667/rev ..

Align seg 1/1 to reverse of: AL539667 from: 1 to: 847

266 GlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgG1 282

|||||
846 GGCTGCAAGACGCGCTGCAGCTKTTGATCGAGACTCCTCGAGAGGGG 797
282 yGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuP 299
|||||
796 AGAGCGGCTGGACATGCGAGCCACTAAAGCAATCTTCAACCGAACTCCTCT 747
299 heGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyr 315
|||||
746 TTGGAGGACACGAACCAAGCGCCAGTGCAGCCACATCTCTGATCATTAC 697
316 leuGlyLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSe 332
|||||
696 CTGGGGCTCCACCCACACGCTCCCCAGAAAGCGCGAGAAGAGCTGAAGAG 647
332 rLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluT 349
|||||
646 TAAGGTTTACTTTGCAAGMCAATCAAGACACAAAGTTGGACATGGAAA 597
349 hrLeuGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArg 365
|||||
596 TTTTGAACAACATCMAATMCCNCGGGTGTGTATTAAAGAGACCCCTCGA 547
366 leuAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheG1 382
|||||
546 CTGAACCCCCMGNCCAGAGAGGTTNCGGNGNCTCCGAASMCTTCTGA 497
382 uLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSerIleC 399
|||||
496 ATCAATGATGATCCMGANCCCAAGGGCTGGAATGTATCTACAGATATCT 447
399 ysAspThrHisAspValAlaAspIlePheThrAsnLysGluPheAsn 415
|||||
446 GTGATACCCATGATGTGGCAGAGATCTTCAACAACAAGAAGATTTAAC 397
416 ProAspArgPheIleValProHisProGlnAspAlaSerArgPheSerPh 432
|||||
396 CCTGACCGATGCTGCTGCCTCACCCMGAGACGCACCCMGGCTCAGCCC 347
432 elleProPheGlyGlyLeuArgSerCysValGlyLysGluPheAlaL 449
|||||
346 CATCCCATTTGGAGAGAGCCCTTAGGAGCTGTGTAGGCCAAAGAATTGCGMA 297
449 ySileLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrp 465
|||||
296 AAATTCCTCCCAAAATATTATTACAGTGGAGCTGSCCAGGCATTTGTGACTSG 247
466 GlnLeuLeuAsnGlyProProThrMetLysThrSerProThrValTyrPr 482
|||||
246 CCSCBCTTAATGGAACCTCTACAAABGAAAAACCAAGNCCACCGTGTATCC 197
482 oValAspAsnLeuProAlaArgPheThrTyrPheGlnGlyAsp 496
|||||
196 TGTGACAATCTCCTGCAAGATTCAACCATTTCCCATGGGAA 154
seq_name: gb_est48:AW513600

seq_documentation_block: 618 bp mRNA EST 03-MAR-2000
LOCUS AW513600
DEFINITION xc47D06.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2707091 3'
similar to SW:CP26_HUMAN O43174 CYTOCHROME P450 26 ;, mRNA
sequence.

ACCESSION AW513600
VERSION AW513600.1 GI:7151678
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 618)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
Image.llnl.gov/image/html/iresources.shtml

Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco
High quality sequence stop: 440.
Location/Qualifiers

FEATURES
source
1. 618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2707091"
/clone_lib="NCI-CGAP_Utl1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"

BASE COUNT 111 a 191 c 199 g 115 t 2 others
ORIGIN

alignment_scores:
Quality: 982.00 Length: 206
Ratio: 4.935 Gaps: 0
Percent Similarity: 96.602 Percent Identity: 92.233

alignment_block:
US-09-668-482-32 x AW513600 ..

Align seg 1/1 to: AW513600 from: 1 to: 618

34 SerSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetG1 50
|||||
1 AGCGGGCCGCGACCGAGTGTGCTCCCTCCATTGCCCGCGGACTATGGG 50
50 ypheProphneheglygluThrLeuGlnMetValLeuGlnArgArgLysp 67
|||||
51 CTTCCCTCTTTGGGAAACCTTGACATGTACTGCAGCGGAGGAAGT 100
67 heLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 83
|||||
101 TTCCTGACAGTAAAGCGACGAGAAATACGGCTTCACTACAAGACGCATCTG 150
84 pheGlyArgProThrValArgValMetGlyAlaAspAsnValArgArg1 100
|||||
151 TTCGGGCGGCCACCGTACGGGTGATGGCGCGGACAATGTGGCGGCAT 200
100 eleuLeuGlyLuhHisArgLeuValSerValHisTrpProAlaSerValA 117
|||||
201 CTTCGCTCGAGACACCGGCTGTGCTGCCACTGGCCAGCGTGGTGC 250
117 rgThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHis 133
|||||
251 GCACCATCTCGGATCTGGCTGCCCTCTTAACCTGCACGACTCTCGCAC 300
134 LysGlnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuG1 150
|||||
301 AAGCAGCGCAAGAGGATGATTGCGGGCTTCAGCCGCGAGGACTCGA 350
150 ncystYrValLeuValIleAlaGluGluValSerSerCysLeuGluGlnT 167
:|||||
351 ATGCTACGTGCGGTGATCACCGAGGAAGTGCGCACAGACCTGGAGCAGT 400

167 rpleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183
|||||
401 GGCTGAGCTGCGCGAGCGCGGCTCTGCTCTACCCCGAGGTGAAGCCG 450
184 leuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPr 200
|||||
451 CTCATGTTCCGAATCGCCATGCGCATCTTGTGCTGCGGAACCCAACT 500
200 oAlaGlyGlyGlyLuhAspGluGlnGlnLeuValGluAlaPheGluGlu 217
|||||
501 GCGCGGCGACGCGGACTCCGAGACAGACGCTTGTGAGAGGCTTCGAGGAAA 550
217 etThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
|||||
551 TGACCCGCAATCTCTCTCGCTGCCCATCGACGTGCNCTTCAGCGGCGCTG 600
234 TyrArgGlyValLysAla 239
|||||
601 TACCGGGGCATGAAGGCN 618

seq_name: gb_est72:BE236243

seq_documentation_block:
LOCUS BE236243 537 bp mRNA EST 10-JUL-2000
DEFINITION 143959 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE236243
VERSION BE236243.1 GI:9020961
KEYWORDS EST.
SOURCE
ORGANISM
Sus scrofa
pig.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 537)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
TITLE
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL
Unpublished (2000)
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 85 row: B column: 21
Seq primer: ATTTCGTGACACTATAG.

FEATURES
source
1. 537
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

BASE COUNT 96 a 163 c 174 g 104 t
ORIGIN

alignment_scores:
Quality: 900.00 Length: 179
Ratio: 5.114 Gaps: 0
Percent Similarity: 98.324 Percent Identity: 94.972

alignment_block:

US-09-668-482-32 X BE236243

Align seg 1/1 to: BE236243 from: 1 to: 537

25	LeuLysLeuThrPaspLeuTyrCysValSerSerAlaGspArgSerCysAl	41
:		:
1	ATCAAACTCTGGGACCCTGTACTGCGTGAGCAGCCGGAGACCGAGCTGCAC	50
41	AlaLeuProLeuProProGlyThrMetGlyPheProPhePheGlyLuhrl	58
:		:
51	CCCTTCCTTTGGCCCCCTGGAACTATGGCTTCCCTTCTTTGGGGAGACAT	100
58	euglMetValLeuGlInArgArglyspheLeuGlMetLysArgArglys	74
:		:
101	TGCAGATGCTGTACAGCGGAAGAGTCTCTGCAGATGAAGCGCAGAGAAA	150
75	TyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrValArgVa	91
:		:
151	TACGGTTTCATCTACAAGACGCACTGTTCGGAGAGCCACCGTGGGGT	200
91	ImetClyAlaAspAsnValArgArgIleLeuLeuGlyGluHisArgLeuV	108
:		:
201	GATGGGTGCAGACACAGCTGGCGGCATCTTGCTCGGGGAACCCGGCTCG	250
108	AlSerValHisTrpProAlaSerValArgThrIleLeuGlyAlaGlyCys	124
:		:
251	TGTCGGTCCACTGGCCGGCGTGGTGCGCAGATCCTGGCGTCTGCGCTGC	300
125	LeuSerAsnLeuHisAspSerSerHisLysGlnArgLysValIleMe	141
:		:
301	CTCTCCAACTGCACGACTCCTTCGCACACAAGCAGCGCAAGAGTGATTAT	350
141	tglnAlaPheSerArgGluAlaLeuGlnCysTyrValLeuValIleAlaG	158
:		:
351	GCAGGCCCTTCAGCCCGGAGGCGCTCCAGTGTACTAGTGGCGGTGATCGCAG	400
158	IuGluValSerSerCysLeuGlnGlnTrpLeuSerCysGlyGluArgGly	174
:		:
401	AGGAAGTGGACAGTTCCTGAGCAGCAGTGGCTGAGCTGCGGAGAGCGCGGC	450
175	LeuLeuValTyrProGluValLysArgLeuMetPheArgIleAlaMetAr	191
:		:
451	CTCCTGGTTTACCCCCAGGTGAACGCCCTCATGTTCGCGCATCGCATGCG	500
191	gileLeuLeuGlyCysGluProGlyProAlaGlyGly	203
:		:
501	CATCCTGCTGGGCTGCGAGCCCCGGCTGGGAGCGCG	537

seq_name: gjb_est87:BF385034

seq_documentation_block:

LOCUS	BF385034	539 bp	mRNA	EST	27-NOV-2000
DEFINITION	602045632F1	NCI_CGAP_L19	Mus musculus	cdna clone	IMAGE:4195422 5'

ACCESSION BF385034

KEYWORDS EST.

ORGANISM Mus musculus

Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus

AUTHORS AND **EXPERIENCE** 1 (bases 1 to 333)

NIH-MGC <http://mgc.ncl.nih.gov/>

THE NATIONAL INSTITUTE OF HEALTH, MAMMALIAN GENE COLLECTION (MGC) JOURNAL
Published (1999)

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: csraeb@r@gmail.com

Tissue Procurement: Jeffrey E. Green, M.D.,
CDNA Tissue Procurement, 1155 E. 1st St.,
Seattle, WA 98101

CDNA Library Arrayed by: The I.M.A.G.E. Consortium

Clone distribution: MGC clone distribution information ca

found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L1AM9530 row: 0 column: 07
 High quality sequence stop: 539.

FEATURES	Location/Qualifiers
source	1. .539

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4195422"
/clone_lib="NCI_CGAP_L19"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

```

BASE COUNT	86 a	160 c	181 g	112 t
ORIGIN				

alignment_scores:

Quality:	888.00	Length:	171
Ratio:	5.224	Gaps:	0
Percent Similarity:	99.415	Percent Identity:	99.415

alignment_block:

US-09-668-482-32 x BF385034

Align seg 1/1 to: BF385034 from: 1 to: 539

1	MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr	17
25	ATGGGGCTCCCGGCGCTGCTGGCCAGTGGCTCTGCACCTTGTGCTGCC	74
17	oLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS	34
75	GCTGCTGCTCTTCTCTGGCGGCGCTCAAGCTCTGGGACCTGTACTGTGTA	124
34	erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly	50
125	GCAGCCGCGATCGCAGCTGGCGCCCTCCCTTGCCCCCGGTACCATGGGC	174
51	PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh	67
175	TTCCCATTTCTTTGGGAAACATGTGACAGATGGTCTTACGCGAGAGATT	224
67	eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP	84
225	TCTGCAGATGAAGCGCAGAAATACGGCTTCATCTACAAAGACGATCTGT	274
84	heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle	100
275	TTGGGGGGCCACACGGTGGGTGATGGGGCGGATATATGTGGCGGCATC	324
101	LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr	117
325	TTGCTGGGAGAGACACCGGTGGTGTGGTGCACCTGAGCCCGCTGGTGG	374
117	gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL	134
375	CACCATCTCGGGCGCTGGCTGCCCTTCCAACTGCACGATTCCTTCGCACA	424
134	ysGlnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln	150
425	AGCAGCGAAGAAGGTGATTATGACAGGCCCTTCAGCCGCGAGGACATCCAG	474
151	CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGluGlnTr	167
475	TGCTACGTGCCCGTGAATCGCTGAGGAAGTACAGCAGTTGTCTGAGCAGTG	524
167	pleuSerCysGly	171
525	GCTAAGCTGCGGGC	537

```
seq_name: gb_est89:BF533769

seq_documentation_block:
LOCUS      BF533769      564 bp      mRNA
DEFINITION 602073982F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4210893 5',
            mRNA sequence.
ACCESSION  BF533769
VERSION    BF533769.1  GI:11621132
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 564)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: L1AM9779 row: c column: 22
            High quality sequence stop: 564.

FEATURES
    source
        1..564
            /location=Qualifiers
            /organism="Mus musculus"
            /strain="FVB/N"
            /db_xref="taxon:10090"
            /clone="IMAGE:4210893"
            /clone_lib="NCI_CGAP_L19"
            /lab_host="DH10B (T1 phage-resistant)"
            /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.9 kb. Constructed by Life
            Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT      89 a      172 c      189 g      114 t
ORIGIN
alignment_scores:
    Quality:      888.00      Length:      171
    Ratio:        5.224      Gaps:      0
    Percent Similarity: 99.415      Percent Identity: 99.415

alignment_block:
US-09-668-482-32 x BF533769 ..
Align seg 1/1 to: BF533769 from: 1 to: 564

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
50 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCTCTGCACCTTCGTGCTGCC 99
|||||
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTirPaspLeuTyrCysVal 34
|||||
100 GCTGCTGCTCTTCCTGCGCGGCTCAAGCTCTGGACCTGTACTGTGTGA 149
|||||
34 erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
150 GCAGCCCGCATCGCAGCTGCGCCCTCCCTGCCCCCGGTACCATGGGC 199
|||||
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgGlySph 67
|||||
200 TTCCCATTTCTTTGGGAACATTCAGATGTGCTTCAGCGGAGGAGTT 249
|||||
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
```

```
250 TCTGCAGATGAAGCCGACGAATAATACGGCTTCATCTTACAGACGCACTCTGT 299
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
300 TTGGGGCGGCCACCGGTGCGGGTGATGGGCGCGGATATGTGCGCGCATC 349
|||||
101 leuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
350 TTGCTGGAGAGACACCGGTTGGTGTCCGTGCACCTGCCCCGCTGGTGGC 399
|||||
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
400 CACCATCCTGGCGCTGGCTGCTCTCCAACTGCACGATTCCTCGCACA 449
|||||
134 ysgLArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
|||||
450 AGCAGCGAAAGAGGTGATTATGCAGGCCCTTCAGCCGCGAGGCACTCCAG 499
|||||
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGlnTr 167
|||||
500 TGCTACGTGCCCGGTGATCGCTGAGGAAGTCAGCAGATTGTCTGAGCAGTGT 549
|||||
167 pleuSerCysGly 171
|||||
550 GCTAAGCTGCGGC 562

seq_name: gb_est71:BE189825

seq_documentation_block:
LOCUS      BE189825      696 bp      mRNA
DEFINITION db61c05.y1 wellcome CRC psk egg Xenopus laevis cDNA clone
            IMAGE:3377480 5' similar to gb:gb|AF057566.1|AF057566 Xenopus
            laevis retinoid acid converting enzyme (XENOPUS);, mRNA sequence.
ACCESSION  BE189825
VERSION    BE189825.2  GI:9729548
KEYWORDS   EST.
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 696)
AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
            WashU Xenopus EST project, 1999
            Unpublished (1999)
            On Jun 22, 2000 this sequence version replaced gi:8668718.
            Other_ESTs: db61c05.x1
            Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B.
            Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
            University Genome Sequencing Center
            Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 483.

FEATURES
    source
        1..696
            /location=Qualifiers
            /organism="Xenopus laevis"
            /db_xref="taxon:8355"
            /clone="IMAGE:3377480"
            /clone_lib="Wellcome CRC psk egg"
            /tissue_type="egg"
            /lab_host="DH10B (phage-resistant)"
            /note="Vector: pBluescript SK-, Site_1: NotI; Site_2:
```

ECORI; cDNAs were oligo-dT primed and directionally cloned. Library was constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B. Guirion (Wellcome/CRC Institute).
Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT 156 a 200 c 202 g 138 t
ORIGIN

alignment_scores: Quality: 851.00 Length: 235
 Ratio: 4.234 Gaps: 3
Percent Similarity: 85.532 Percent Identity: 71.915

alignment_block:
US-09-668-482-32 x BE189825 ..

Align seg 1/1 to: BE189825 from: 1 to: 696

```
43 ProLeuProProGlyThrMetGlyPheProPheGlyGluThrLeuG1 59
|||||
8 CCACTGCCCCCGGCACTATGGGGCTGCCCTTCTTCGGGAGACTCTGA 57
59 nMetValLeuGlnArgArgLysPheLeuGlnMetLysArgArgLysTyrg 76
|||||
58 AATGCTGCTGCAGAGGCGCAAGTTCCTCAATGAAGCGTAGAAAGTACG 107
76 lypHeileTyrlYstThrlHisleupheGlyArgProThrValArgValMet 92
|||||
108 GTGCACTACAAAGACGACATCTGTTCGTAGCCACGCGTGCAGCAG 157
93 GlyAlaAspAsnValArgArgileLeuLeuGlyGluHisArgLeuValSe 109
|||||
158 GGGCAGAGAAAGCTTCGCCAGATCCCTATTGGGGAGACACACAGCTGGTTC 207
109 rValHisTrpProAlaSerValArgThrileLeuGlyAlaGlyCysLeus 126
|||||
208 GGTGCACTGGCCGGCCCTCGGTGCGCACGATCCTAGGGCCGGCTGTCTGT 257
126 eAsnLeuHisAspSerSerHisLysGlnArgLysLysValileMetGln 142
|||||
258 CCAACCTGCACGACTCTGTAGCACACACCAAGAAAGTATTGCACAA 307
143 AlaPheSerArgGluAlaLeuGlnCysTyrlValLeuValileAlaGluG1 159
|||||
308 GCCTTCTCCCGAGATGCCCTCGACAATTACGTCCGCGAGATGGAAGAAGA 357
159 uValSerSerCysLeuGlnTrpLeuSerCysGlyGluArgGlyLeuL 176
|||||
358 GGTGAGACGCTCTGTAAACCTGTGCTGCAGAGCGGCCCC...TGCGTGC 404
176 euValTyrlProGluValLysArgLeuMetPheArgileAlaMetArgile 192
|||||
405 TGATGTATCCCGCATCAAGCGACTGATGTCCGATTCGCATGAAGCTC 454
193 leuLeuGlyCysGluProGlyProAlaGlyGlyGluAspGluGlnG1 209
|||||
455 CTGCTCGGCTGCGATCCC...CAGCGCATGGACAGCCACACATGAGAGAC 501
209 nLeuValGluAlaPheGluGluMetThrArgAsnLeuPheSerLeuProI 226
|||||
502 GCTGCTCGAAGCCTTCGAGAAATGACCCGAAATCTCTTCTGTGCTTA 551
226 leaspValProPheSerGlyLeuTyrlArgGlyValLysAlaArgAsnLeu 242
|||||
552 TTGATGTGCCATGTAGCGGCTCTACCGGGGTCTGCGGCTAGAGACCTT 601
243 lHeHisAlaArgileGluGluAsnileArgAlaLysileArgArgLeuG1 259
|||||
602 ATTCAATCGCGAATTGATGAACAATCGAAGAGAAAGCTGCTAAGG.... 646
259 nAlaThrGluProAspGlyGlyCysLysAspAlaLeuGlnLeuLeuileG 276
|||||
647 .....GAACCAGACGATTAATTGCCGGGATGCCCTGCAGCTGCTGATTG 689
```

276 luhis 277
:::
690 ACTAT 694

seq_name: gb_est4:AA239785

seq_documentation_block:
LOCUS AA239785 474 bp mRNA EST 03-MAR-1997
DEFINITION mx80a03.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:692620 5'
similar to TR:G1001252 G1001252 HYPOTHETICAL 50.6 KD PROTEIN ;
mRNA sequence.

ACCESSION AA239785 GI:1863825
VERSION AA239785.1 GI:1863825
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 474)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:426180

Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 461.

FEATURES
source

1..474
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:692620"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGGAGGCGCGCAATCTTTTCTTTTCTTTTCTTTT
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT 135 a 99 c 118 g 122 t
ORIGIN

alignment_scores: Quality: 845.00 Length: 158
 Ratio: 5.382 Gaps: 0
Percent Similarity: 99.367 Percent Identity: 99.367

alignment_block:
US-09-668-482-32 x AA239785 ..

Align seg 1/1 to: AA239785 from: 1 to: 474

```
324 GlnLysValArgGluGluLileLysSerLysGlyLeuLeuCysLysSerAs 340
|||||
1 CAGAAAGTTCGAGAAGAGATTAAGAAGCAAGGCGTCTTACCTTGCAGAGCAA 50
```


340 nglInaspasnlysleuAspmetGluThrleuGluInleuLysTyrIleG 357
|||||
51 TCAAGACAACAAGTTAGACATGGAACCTTTGGAACAGCTTAATACACTG 100
357 lYcysValIleLysGluThrleuArgleuAsnProProValProGlyGly 373
|||||
101 GGTGTGTCATTAAAGAGACCCCTGCATTGAATCCTCCGGTTCAGAGAGG 150
374 PheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLy 390
|||||
151 TTTCGGGTGCTCTGAAGACTTTTGAGCTGAATGATACAGATCCCAA 200
390 sGlyTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaAspI 407
|||||
201 GGGCTGGAATGTATTATTACAGTATCTGTACACCCACGATGTGCAGATA 250
407 lPheThrAsnLysGluGluPheAsnProAspArgPheIleValProHis 423
|||||
251 TCTTCACTAACAAAGGAGGAATTAATCCGACCCGCTTATAGTCCCTCAT 300
424 ProGluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuAr 440
|||||
301 CCAGAGGATGCTTCCCGGTTCCAGCTTCATTCCATTGGAGAGGCCCTCG 350
440 gSerCysValGlyLysGluPheAlaLysIleLeuLysIlePheThrV 457
|||||
351 GAGCTGTGTAGGCCAAAGAGTTTGCAAAATCTCTTAAGATATTTTACAG 400
457 aGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThr 473
|||||
401 TGGAGCTGGCTAGGCACTGTGATGTGCAGCTTCTAATGACCTCCTACA 450
474 MetLysThrSerProThrValTyr 481
|||||
451 ATGAAGACAAGCCCCCACTGTGTAC 474
seq_name: gb_est85:BF236872
seq_documentation_block: 545 bp mRNA EST 14-NOV-2000
LOCUS BF236872 602027980F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4163202 5',
DEFINITION mRNA sequence.
ACCESSION BF236872
VERSION BF236872.1 GI:11150789
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 545)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9446 row: P column: 19
High quality sequence start: 2
High quality sequence stop: 545.
Location/Qualifiers
1. 545
FEATURES
source
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4163202"
/clone_lib="NCI_CGAP_L19"

/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 87 a 165 c 184 g 109 t
ORIGIN
alignment_scores:
Quality: 805.00 Length: 173
Ratio: 4.820 Gaps: 4
Percent Similarity: 96.532 Percent Identity: 94.220
alignment_block:
US-09-668-482-32 x BF236872 ..
Align seg 1/1 to: BF236872 from: 1 to: 545
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
35 ATGGGGCTCCCGCGCTGCTGCCAGTGGCGCTCTGCACCTTCGTGCTGCC 84
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspleuTyrCysValS 34
|||||
85 GCTGCTGCTCTTCCGTGGCGGCGCTCAAGCTCTGGACCTGTACTGTGTGA 134
34 erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
135 GCAGCCGCGATCGCAGCTGCCCTCCCTTGCCCCCGGTACCATGGGC 184
51 PheProPhe.PheGlyGluThrleuGlnMetValLeuGlnArgArgLysP 67
|||||
185 TTCCCATCTCTTGGGGA..AACATGCAGATGCTGCTCAGCGGAGAGAAGT 231
67 heLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 83
|||||
232 TTCTGCAGATGAAGCGCAGCAAAATACGGCTTCATCTACAGACGCATCTG 281
84 PheGly.ArgProThrValArgValMetGlyAlaAspAsnValArgArgI 100
:::||||
282 TTGGGGCGGCC..ACGGTGGGGGTGATGGCGCGGAGATAATGTGGCGGCA 328
100 leLeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerVal 116
|||||
329 TCTTGTGGGAGAGACCGCTTGCTGTCGGTGACATGCCGCCGCTCGGTG 378
117 ArgThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHi 133
|||||
379 CGCACCATCTCGGGCGCTGGCTGCCCTCCAACCTGCACGATTCCTCGCA 428
133 sLysGlnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuG 150
|||||
429 CAAGCAGCGAAGAAGGTGATTATGCAGAGGCTTCAGCGCGAGGCACTCC 478
150 IncysTyrValLeuValIleAlaGluGluValSerCysLeuGluGln 166
|||||
479 AGTGCTACGTGCCCGTGAATCGCTGAGGAAGTCAAGCAGATTGTCTGGAGCAG 528
167 TrpLeuSerCysGly 171
|||||
529 TGGCTAAAGCTGCGGC 543
seq_name: gb_est82:BF055367
seq_documentation_block: 646 bp mRNA EST 16-OCT-2000
LOCUS BF055367 7378f12.x1 Soares_NSF_F8_9W_OT_PA_P.S1 Homo sapiens cDNA clone
DEFINITION IMAGE:3392591 3' similar to SW:CP26_HUMAN O43174 CYTOCHROME P450 26
ACCESSION BF055367
VERSION BF055367.1 GI:10809263
KEYWORDS EST.
/: mRNA sequence.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 646)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Glibco
High quality sequence stop: 450.
FEATURES
source
1. 646
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3392591"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and cloneids: Soares NbHSF pool 1: 309384-310919, 323208-325895 Soares Nb2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares Nb2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NbHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NbHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 202 a 129 c 120 g 194 t 1 others
ORIGIN
alignment_scores:
Quality: 743.00 Length: 148
Ratio: 5.160 Gaps: 0
Percent Similarity: 97.297 Percent Identity: 91.216
alignment_block:
US-09-668-482-32 x BF055367/rev ..
Align seg 1/1 to reverse of: BF055367 from: 1 to: 646
350 LeuGlulnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgIle 366
||| ::|||
643 TTGGCCCACTTAATACTCTCGGGTGTATTAAAGAGACCCCTTNGACT 594
366 uAsnProValProGlyGlyPheArgValAlaLeuLysThrPheGluL 383
|||||
593 GAATCCCCAGTTCAGAGAGGGTTTCGGGTCTCTGAAGACTTTTAATT 544
383 euAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCys 399
|||||
543 AAAATGATACACAGATTCCCAAGGGCTGGAATGTATCTACAGTATCTGT 494
400 AspThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnPr 416
|||||
493 GATACTCATGATGTGGCAGAGATCTTCACCAACAAGAGAATTAAATCC 444
416 oAspArgPheIleValProHisProGluAspAlaSerArgPheSerPheI 433
|||||
443 TGACCGATTCATGCTGCCTCACCCACAGAGATGCATTCAGGTTACGTTCA 394
433 leProPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLys 449

|||||
393 TTCCATTTGGAGAGGCGCTTAGAGCTGTGTAGCGCAAGAAATTTGCCAAA 344
450 IleLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpG1 466
|||||
343 ATTCTTCTCAAAATATTTAACAGTGGAGCTGGCCAGGCATTTGTGACTGGCA 294
466 nLeuLeuAsnGlyProProThrMetLysThrSerProThrValTyrProV 483
|||||
293 GCTTCTAAATGGACCTCTACAAATGAAACCAGTCCACCGTGTATCCTG 244
483 aLAspAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||
243 TGGACAATCTCCCTGCAGAGATTCCACCATTTCCATGGGGAATC 200
seq_name: gb_est43:AW174347
seq_documentation_block:
LOCUS AW174347 .720 bp mRNA EST 16-NOV-1999
DEFINITION f142a06.y1 Sugano Kawakami zebrafish DNA Danio rerio cDNA clone 2640274 5' similar to SW:CP26_BRARE P79739 CYTOCHROME P450 26 ; , mRNA sequence.
ACCESSION AW174347
VERSION AW174347.1 GI:6440295
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 720)
AUTHORS Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S., Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T., Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU zebrafish EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: f142a06.x1
Contact: S.L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by: Washington University Genome Sequencing Center
Seq primer: T3 ET from Amersham
High quality sequence stop: 478.
FEATURES
source
Location/Qualifiers
1. .720
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="2640274"
/clone_lib="Sugano Kawakami zebrafish DNA"
/sex="mixed (one male and one female, including unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGACACA."
BASE COUNT 172 a 176 c 196 g 176 t

ORIGIN

alignment_scores: Quality: 742.50 Length: 238
 Ratio: 3.750 Gaps: 3
Percent Similarity: 83.193 Percent Identity: 62.185

alignment_block:
US-09-668-482-32 x AW174347 ..

Align seg 1/1 to: AW174347 from: 1 to: 720

```
1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||       :::|:::|:::|:::|:::|:::|:::|:::|:::|
25 ATGGGGCTGTACACCCCTTATGGTCACCGTTGCTGTCCATCGTGTACC 74
17 OleuleuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
75 CGTTTACTCTTCTCGCCGCGGTGAAGTTGTGGAGATGTTAATGATCC 124
34 eSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||       :::|:::|       |||||       |||||       |||||
125 GACGAGTCGATCCGAACTGCAGAGTCTCTACCGCAGGTACCATGGGC 174
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgGlySph 67
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
175 TTGCCGTTCAATTGGAGAAACGCTCCAGCTGATCCTCCAGAGAGAAGT 224
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||       :::|:::|       |||||       |||||       |||||
225 TCTGCCGATGAAGCGCAGAAATACGGGTGCAATCCTCAAGACCGACCTCT 274
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||       :::|:::|       |||||       |||||       |||||
275 TCGGGAACCCGACTGTCAAGGGTGAGGCTGATATATGTGAGGCAGATT 324
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||       :::|:::|       |||||       |||||       |||||
325 CTGCTGGGCGAACAAGCTGTGTCTGTTCAGTGGCCAGCATCATGAGAG 374
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||       :::|:::|       |||||       |||||       |||||
375 AACCATCCTGGGCTCTGACACACCCTCTCCAATGTCCATGAGTCAACACA 424
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
425 AAACACAGAAAGGCCATATATGAGGCGCTTCTCTCGAGATGCTCTGAG 474
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGluInTr 167
|:::|       :::|:::|       :::|:::|       :::|:::|       :::|:::|
475 CACTACATTCCTCGATCCATCAGAGGTGAAGCGCCATACAGGAATG 524
167 pLeu.....SerCysGlyGluArgGlyLeuLeuValTyrProGluV 181
|||||       |||||       :::|:::|       :::|:::|       :::|:::|
525 GCTGCACAAAAGACTCCTGC.....GTGCTGGTTTATCCAGATA 562
181 alLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
:::|:::|       :::|:::|       :::|:::|       :::|:::|       :::|:::|
563 TGAAGATACTCATGTTCGGGATAGCTATGAGAATCCTGCTGTGTTGAA 612
198 ProGlyProAlaGlyGlyGluAspGluGlnGlnLeuValGluAlaPh 214
|||       |||       :::|:::|       :::|:::|       :::|:::|
613 CCACAGCAATA.....ATGACTGACGAGCATGAGCTGTGGAAGCTTT 656
214 eGluGluMetThrArgAsnLeuPheSerLeuProIleAspValProPhes 231
:::|:::|       :::|:::|       :::|:::|       :::|:::|       :::|:::|
657 GGAGCGAATGATCAGCGAGCTGTGCTGTGCAATCCACGTTCTTTCA 706
231 erGlyLeuTyrArg 235
|||       |||       |||       |||       |||       |||
707 TTGGTCTGTACAGG 720
```

seq_name: gb_est51:AW765767

seq_documentation_block:

LOCUS AW765767 669 bp mRNA EST 16-FEB-2001
DEFINITION da77a02.y1 Harland stage 19-23 Xenopus laevis cDNA clone
IMAGE:3200906 5'-similar to gb:gb|AF057566.1|AF057566 Xenopus
laevis retinoic acid converting enzyme (XENOPUS);, mRNA sequence.

ACCESSION AW765767
VERSION AW765767.1 GI:7697743

KEYWORDS EST.
SOURCE African clawed frog.

ORGANISM

Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

REFERENCE

AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Page,D.
 ', Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
 ', B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
 Waterston,R. and Wilson,R.
 Washu Xenopus EST project, 1999

TITLE

JOURNAL Unpublished (1999)
COMMENT Other_ESTs: da77a02.x1

Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Library constructed by R. Harland, PhD.(University of California,
Berkeley)

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LNL at:
image.lnl.gov/image/html/lireources.shtml
Seq primer: -40RP from Gibco
High quality sequence stop: 495.

FEATURES

source

location/Qualifiers
1..669
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:3200906"
 /clone_lib="Harland stage 19-23"
 /tissue_type="neurula"
 /dev_stage="stage 19-23"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: PCS107 (custom); Site_1: NotI; Site_2: SalI
 ; cDNA made by oligo-dT priming. Library constructed by
 Dr. Francesca Mariani in the laboratory of R. Harland,
 Ph.D. (University of California, Berkeley). References:
 XBF-2 is a transcriptional repressor that converts
 ectoderm into neural tissue. Mariani, FV, Harland, RM.,
 Development. 1998 Dec;125(24):5019-31. PMID: 9811586; UI:
 99030283; Use of large-scale expression cloning screens in
 the xenopus laevis tadpole to identify gene function.
 Grammer TC, Liu KJ, Mariani FV, Harland RM., Dev Biol.
 2000 Dec 15;228(2):197-210. PMID: 11112324; UI: 20564075;
 Note: This is a Xenopus Gene Collection (Xgc) library."

BASE COUNT 137 a 206 c 187 g 139 t
ORIGIN

alignment_scores: Quality: 736.50 Length: 207
 Ratio: 4.209 Gaps: 2
Percent Similarity: 84.541 Percent Identity: 70.531

alignment_block:
US-09-668-482-32 x AW765767 ..

Align seg 1/1 to: AW765767 from: 1 to: 669

seq_name: gb_est23:A167038

seq_documentation_block:

LOCUS A1667038 640 bp mRNA EST 18-MAY-1999
DEFINITION fc24h03.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
SW:CP26_BRARE P79739 CYTOCHROME P450 26 ;, mRNA sequence.
ACCESSION A1667038
VERSION A1667038.1 GI:4805394
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.

REFERENCE
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Jackson,X., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.

```

TITLE      WashU Zebrafish EST Project 1998
JOURNAL    Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenzentrumPrimardatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Possible reversed clone: similarity on wrong strand
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 474.

FEATURES
Source
1..640
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'GACTAGTTCTAGATCGCAGCGGCCGGCCCTTTT'TTTT'TT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."
BASE COUNT   156 a       162 c       167 g       155 t
ORIGIN
alignment_scores:
Quality:      683.50          Length:      197
Ratio:        3.974           Gaps:         2
Percent Similarity: 87.310    Percent Identity: 67.005

alignment_block:
US-09-668-482-32 x AI667038 ..

Align seg 1/1 to: AI667038 from: 1 to: 640

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysTrpPheValLeuPr 17
||||| :|||:::||||| |||||:::|||||
62 ATGGGGGTGTACACCCTTATGTCACCTTCTCTGCACCATCGTGCTACC 111
17 OleuleuLeuPheleuAlaAlaleuLysleuTrpAspleuTyrcysVals 34
|:::|||||:::|||||:::|||||:::|||||:::
112 CGTTTACTCTTTCTCGCCGCGGTGAAGTTGTGGAGATGTTAATGATCC 161
34 erSerArgaspargserCysAlaLeuProleuProProglyThrMetgly 50
||| ::||| | ||||| ||||| ||||| |||||
162 GACGAGTCGATCCGAACTGCAGAAGTCCCTTACCGCCAGGTACCATGGC 211
51 PheProPhePheglVgIuThrlengInmetValIengInArDrrtlwspH 67

```



```

:::|||||:::|||||:::|||||:::|||||
212 TTGCCGTTCAATGGAGAAACGCTCCAGCTGATCCTCCAGAGAGAGATT 261

67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 84
   |||:::|||||:::|||||:::|||||
262 TCTGCGCATGAAACGGCAGAAATACGGGTGCATCTACAAGACGCACTCT 311

84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
   |||:::|||||:::|||||:::|||||
312 TCGGGAACCCGACTGTCAAGGTGATGGAGCTGATATGTGAGGCAGATT 361

101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
   |||:::|||||:::|||||:::|||||
362 CTGCTGGGCGAACAACAAGCTGGTGTCTGTCAGTGGCCAGCATCAGTGAG 411

117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
   |||:::|||||:::|||||:::|||||
412 AACCATCTGGGCTCTGACACCCCTCTCCAATGTCCATGGAGTTCAACACA 461

134 ySGlnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
   |||:::|||||:::|||||:::|||||
462 AAAACAAGAAAAAGCCATTATGAGGGCTTCTCTCGAGATGCTCTGGAG 511

151 CysTyrValLeuValIleAlaGluValSerSerCysLeuGluGlnTr 167
   |||:::|||||:::|||||:::|||||
512 CACTACATTCCTCGTGTATCCAGCAGAGGTGAAGACGCCATACAGGAATG 561

167 pLeu.....SerCysGlyGluArgGlyLeuLeuValTyrProGluY 181
   |||:::|||||:::|||||:::|||||
562 GCTGCAAAAAGACTCCTGC.....GTGCTGTTTATCCAGAAA 599

181 alLysArgLeuMetPheArgIleAlaMetArgIleLeuLeu 194
   |||:::|||||:::|||||:::|||||
600 TGAAGAAACTCATGTTCGGATAGCTATGAGAAATCCTGCTT 640
```

